JOURNAL OF APPLIED PHYSICS

SERIALS DEPARTMENT

JOURNAL OF APPLIED PHYSICS

Vol. 25, No. 7 JULY, 1954

In This Issue

		Page
A New Editor	**Robert L. Sproull**	**819**
Special Issue on High Polymer Physics		
Molecular Structure and the Crystallinity of Long-Chain Polymers	C. W. Bunn	**820**
Relationship of First- to Second-Order Transition Temperatures for Crystalline High Polymers	Raymond F. Boyer	**825**
Crystallization Kinetics in High Polymers. I. Bulk Polymers	Leo Mandelkern, Fred A. Quinn, Jr., and Paul J. Flory	**830**
X-Ray and Infrared Studies on the Extent of Crystallization of Polymers	J. B. Nichols	**840**
Infrared Study of the Crystallization of Polyethylene	F. P. Reding and Alexander Brown	**848**
Crystallization in Butadiene-Styrene Copolymers	Lawrence A. Wood	**851**
Regular Contributed Original Research		
Modified Luneberg Lens	A. S. Gutman	**855**
Certain Transmission and Reflection Theorems	V. Twersky	**859**
Steady-State Solution of the Two-Dimensional Diffusion Equation for Transistors	J. S. Schaffner and J. J. Suran	**863**
Diffusion Coefficient, Solubility, and Permeability for Helium in Glass	W. A. Rogers, R. S. Buritz, and D. Alpert	**868**
Asymptotic Evaluation of the Field at a Caustic	Irvin Kay and Joseph B. Keller	**876**
Secondary Emission from Nichrome V, CuBe, and AgMg Alloy Targets Due to Positive Ion Bombardment	Michael J. Higatsberger, H. L. Demorest, and Alfred O. Nier	**883**
Information Theory and Uncertainty Principle	L. Brillouin	**887**
Decompositions of Inorganic Specimens During Observation in the Electron Microscope	Robert B. Fischer	**894**
Arcing at Electrical Contacts on Closure. Part V. The Cathode Mechanism of Extremely Short Arcs	Paul Kisliuk	**897**
Shadowing Technique for Electron Microscopy—A Possible Substitute for the High-Vacuum Evaporation Technique	Alvin E. Bills and Robert Lefker	**901**
Measurement of the Conductivity of a Jet Flame	Fred P. Adler	**903**
Contact Charging between Nonconductors and Metal	John W. Peterson	**907**
Charge Storage in Junction Diodes	Earl L. Steele	**916**
Instability of Liquid Surfaces and the Formation of Drops	Joseph B. Keller and Ignace Kolodner	**918**
Flicker Noise in Germanium Rectifiers at Very Low- and Audio-Frequencies	D. K. Baker	**922**
Letters to the Editor:		
Low-Loss Ferrites for Applications at 4000 Millicycles per Second	L. G. Van Uitert, J. P. Schafer, and C. L. Hogan	**925**
Continuously Recording Electron Diffraction Camera for Studies of Crystal Structure Transitions	A. Boettcher, R. Thun, and H. Treupel	**926**
Note on the Oxide Replica in Electron Microscopy	Shigeto Yamaguchi	**927**
Second Mode Vibrations of the Pochhammer-Chree Frequency Equation	C. W. Curtis	**928**
Note on the Elastic Wedge Problem	J. H. Huth	**929**
Note on "Effect of Filament Voltage on the Plate Current of a Diode"	Henry S. C. Chen	**929**
Resistance of a Bar with a Transverse Hole	W. R. Smythe	**930**
Erratum: On a Nonlinear Diffusion Equation Applied to the Magnetization of Saturable Reactors	Shou-Hsien Chow	**930**
Index to Advertisers		**xvi**

ROBERT L. SPROULL
Editor

ELMER HUTCHISSON,
EARLE C. GREGG,
Retiring Editors

Board of Editors
MAURICE A. BIOT
RAY E. BOLZ
WALLACE J. ECKERT
ROBERT D. HEIDENREICH
LOUIS A. PIPES
ROMAN SMOLUCHOWSKI
W. R. SMYTHE
ARTHUR V. TOBOLSKY
ALVIN M. WEINBERG

Publication of the
American Institute of Physics

OFFICERS
FREDERICK SEITZ,
Chairman
GEORGE B. PEGRAM,
Treasurer

GOVERNING BOARD
ROBERT F. BACHER
J. W. BUCHTA
KARL K. DARROW
S. A. GOUDSMIT
DEANE B. JUDD
HUGH S. KNOWLES
C. C. LAURITSEN
W. H. MARKWOOD, JR.
WILLIAM F. MEGGERS
PHILIP M. MORSE
BRIAN O'BRIEN
HARRY F. OLSON
GEORGE B. PEGRAM
ERIC RODGERS
RALPH A. SAWYER
FREDERICK SEITZ
WILLIAM SHOCKLEY
RICHARD M. SUTTON
J. H. VAN VLECK
MARK W. ZEMANSKY

Consultant on Printing
MELVIN LOOS

ADMINISTRATIVE STAFF
HENRY A. BARTON,
Director
WALLACE WATERFALL,
Executive Secretary
THEODORE VORBURGER,
Advertising Manager
RUTH F. BRYANS,
Publication Manager
EDITH I. NEFTEL
Circulation Manager
KATHRYN SETZE
Assistant Treasurer

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Robert L. Sproull, Editor, Rockefeller Hall, Cornell University, Ithaca, New York. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of **$8** per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: **$14.00.** Single copies: **$1.25** each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 25, Number 7 July, 1954

UNIVERSITY OF MICHIGAN LIBRARIES

A New Editor—Robert L. Sproull

WITH this issue, a new editor takes over the helm of the *Journal of Applied Physics*. The *Journal* was launched seventeen and a half years ago in an effort to provide a new publication medium devoted particularly to the interests of the applied physicist. The initial stimulus for the new journal came from Mervin J. Kelly, then Director of Research for the Bell Telephone Laboratories and an active participant in the American Institute of Physics' Advisory Council on Applied Physics. Dr. Kelly's suggestion was enthusiastically seconded by the late John T. Tate, who was then Editor of *The Physical Review* and the journal *Physics*, and by Henry A. Barton, Director of the Institute. With the very considerable assistance of Madeline Mitchell Tate, who was then Publication Manager of the Institute, and her staff, the first issue appeared in January, 1937.

To help establish the new journal the American Physical Society turned over the good will and ownership of the journal *Physics* to the American Institute of Physics. The reality of a need for a journal devoted to the interests of applied physicists was then amply demonstrated by the rapid increase in its circulation and the acceptance of the new journal as the official publication medium for many important papers in the applied physics field.

During the years the character of the *Journal of Applied Physics* has changed. Physics is no longer considered by industry as a stepchild to be barely tolerated, as it seemed to be in the early days of the new journal. Instead it has achieved well deserved recognition as the seed from which many new industries are springing and will continue to spring as its potentialities are further realized. During the years of World War II, as this change was taking place, the *Journal* became more of an archive publication and the responsibilities of the Editor both with respect to the Journal and to his other assignments constantly increased. In 1950, Earle C. Gregg became Assistant Editor and aided greatly in the editorial direction of the Journal. It became apparent, however, that if the Journal was to be developed to the fullest extent possible a new editor should be chosen.

Upon the unanimous recommendation of an *ad hoc* committee under the chairmanship of Professor F. V. Hunt, Robert L. Sproull, Associate Professor of Physics at Cornell University, was appointed by the Governing Board of the American Institute of Physics as Editor of the *Journal of Applied Physics*, effective July 1, 1954. Because of the papers already scheduled and the time required for printing operations, Editor Sproull cannot be held wholly responsible for the next few issues. For this reason, the names of Professor Gregg and the writer will be carried on the masthead as retiring editors for the balance of the current year. As Dr. Sproull takes over he carries with him the full confidence and the best wishes of the retiring editors.

ELMER HUTCHISSON

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 7 JULY, 1954

Special Issue on High Polymer Physics

The following six papers were presented at the meeting of the Division of High-Polymer Physics of the American Physical Society in Durham, North Carolina, March 26–28, 1953.

Molecular Structure and the Crystallinity of Long-Chain Polymers*

C. W. Bunn
Imperial Chemical Industries, Ltd., Plastics Division, Welwyn Garden City, England
(Received January 12, 1954)

The molecular characteristics which determine whether a long-chain polymer will crystallize at all—and if so, how readily—are considered. The criterion for crystallization is geometrical regularity of molecular structure; but absolute regularity is not essential, for certain copolymers, as well as certain homopolymers with geometrically irregular structures, do crystallize; the requirement is that the different units which indiscriminately occupy corresponding positions in the structure must be not too different in size and shape (as in monomeric "mixed crystals"); an approximation to regularity is sufficient.

It is unlikely that quite regular molecules can be incapable of crystallizing; when such substances are reported as noncrystalline, it is more likely that they crystallize only with difficulty, and the right conditions have not been found. Reluctance to crystallize is associated with a small interval between relaxation transition ("second-order transition") and melting points; this interval appears to be influenced by the general shape of the molecule (though shape is not the only factor). The temperature range within which crystallization proceeds at an appreciable rate may be still further restricted if the chain repeat distance of the molecules is long; and this same factor may make crystallization at any temperature very slow.

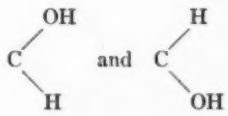
SOME long-chain polymers, notably polythene, crystallize very readily, so much so that it is difficult to obtain them in amorphous glasslike form, even by shock-cooling from the melt; others, like polyethylene terephthalate and polyvinylidene chloride, crystallize well if cooled slowly, but fail to do so (setting instead to a transparent glasslike form) if cooled rapidly; still others, like polystyrene and polymethyl methacrylate, have never been crystallized—at room temperature they are found always in the familiar glasslike condition. What features of molecular structure determine whether a long-chain polymer is capable of crystallizing at all, and if it does whether it will do so readily or only with difficulty? Note that in this inquiry we are not concerned with the melting points (the temperatures of first-order transitions from the crystalline to the amorphous condition), but with the question whether a polymer will crystallize at *any* temperature; natural rubber, for instance is usually noncrystalline at room temperature, but it does crystallize on cooling or on stretching—it is a typical crystallizing polymer, whereas some of the butadiene-styrene copolymers used as "synthetic rubbers" have not been crystallized either by cooling or stretching.

Crystals are regular three-dimensional patterns of atoms, and it is to be expected that a chain polymer will be capable of crystallizing only if its molecular structure is regular, both chemically and stereochemically. This expectation is to a large degree fulfilled by the facts: the typical crystalline polymers are those homopolymers whose molecules are both chemically and geometrically regular in structure—for example, polythene, and the polyesters and polyamides made from straight chain intermediates; in these substances, the same grouping of atoms is repeated regularly along the chain, and there is no possibility of stereochemical irregularity. The occasional branching in polythene does not prevent the crystallization of the sections of molecule between the branch points. Similarly, copolymerization with a minor amount of a second monomer may reduce crystallinity but does not prevent crystal formation altogether, because there are many chain segments containing regular successions of the main constituent, and these can pack together to form crystals. On the other hand, the typical noncrystalline polymers are those in which there is either chemical or stereochemical irregularity or both: many copolymers in which there are comparable amounts of the different monomer constituents are noncrystalline, and among homopolymers it is those in which there are likely to be left- and right-hand groups indiscriminately along the chains which are noncrystalline—for example, polystyrene, polymethyl methacrylate, and polyvinyl acetate, in which the large side groups may occupy left- or right-hand positions at random. Usually, polymers in which a chain carbon atom carries two different substituent atoms or groups (like the three polymers just mentioned) do not crystallize, while those in which a chain carbon atom carries two identical atoms or groups (for example, polyvinylidene chloride $(-CH_2-CCl_2-)_p$) do crystallize.

Absolute regularity of molecular structure is, how-

* Presented at the Symposium on Crystallization of Polymers at the Division of High-Polymer Physics, American Physical Society Meeting, Durham, North Carolina, March 27, 1953.

ever, not essential for crystallization; if there is an approximation to regularity, crystals may be formed. If two substituents on the same chain atom are not too different in size, or if the monomer constituents of a copolymer are not too different in size and shape, crystallization may occur. Polymer crystal structures of this type are analogous to the "mixed crystals" formed from two different nonpolymeric substances if the molecules are similar in shape and size, or in ionic crystals where equivalent sites in the structure are occupied by two different ions indiscriminately, if the ions are of similar size. An example of a polymer crystal of this type is polyvinyl alcohol $(-CH_2-CHOH-)_p$. At one time it was supposed that the crystallinity of this polymer was an indication that all the hydroxyl groups are on the same side of the zigzag plane (at any rate, along the length of chain in any one crystal); but it has now been shown[1] that the hydroxyl groups are randomly placed in left- and right-hand positions. Hydroxyl groups are of course considerably larger than hydrogen atoms, but neither of them are very large compared with the spacings in the crystal, and evidently indiscriminate reversal of the positions of H and OH in the groups



does not lead to strains serious enough to prevent crystallization. The strong hydrogen bonds between the hydroxyl groups of neighboring molecules also help, no doubt, to stabilize the crystal in spite of a certain amount of strain due to misfit.

By contrast, polyvinyl acetate $(-CH_2-CH(O\cdot CO\cdot CH_3)-)_p$ (from which polyvinyl alcohol is made, and which therefore has an irregular structure corresponding to that of the alcohol) does not crystallize. The acetate groups are much larger than hydroxyl groups, and this is evidently the reason for its failure to crystallize. But observe that it is not a matter of mere size in itself—it is a matter of the difference between the sizes of the two substituents on the same carbon atom (in this case $-H$ and $-O\cdot CO\cdot CH_3$). A polymer having two large identical or similar-sized substituents on the same chain atom would have a regular (or approximately regular) structure and would be expected to crystallize, assuming that regularity is the criterion for crystallization. Furthermore, polymers like polyvinyl acetate, polymethyl methacrylate, and polystyrene, if they could be induced by special polymerization conditions to have their substituents in corresponding or regularly alternating space positions, would be expected to crystallize. Molecules with large substituents all in corresponding positions would probably curl into helical configurations (owing to steric effects) and in this case would be quite without symmetry, but lack of symmetry is no bar to crystallization; it is geometrical regularity, not symmetry, which is required for crystallization.

On the question of the size difference between randomly interchanged groups which is tolerable in a crystal structure, there is too little evidence to lead to a general statement. Indeed, it is doubtful whether any general statement will be possible. In ionic mixed crystals, differences (of ionic radii) up to 15 percent can be tolerated; but in polymers the geometrical situation is more complex, and the size limits of substituents on chains will probably depend on the dimensions of the rest of the chain structure, and on the strength of the forces between atoms on neighboring chains. However, in simple polymers some regularities can be discerned. Copolymers of ethylene $CH_2=CH_2$ with tetrafluoroethylene $CF_2=CF_2$, with carbon monoxide and with vinyl alcohol have been found to be crystalline[2]; evidently a CH_2 group (Van der Waals radius of hydrogen 1.1A) may be replaced by a CF_2 group (Radius of F 1.35A) or a CHOH group (a hydroxyl group is somewhat larger than a fluorine atom) or a $C=O$ group (radius of oxygen 1.35A). On the other hand, CHCl or CCl_2 groups are too large (radius of Cl 1.7A), for it is found that heavily chlorinated polythene is noncrystalline. Polyvinyl chloride $(-CH_2-CHCl-)_p$ is interesting in this connection; the amorphous character of chlorinated polythene would lead us to expect polyvinyl chloride to be amorphous also; actually it does show some evidence of crystallinity, but the indications are that such crystallinity as it does possess is due to the presence of a certain proportion of chain segments of regular structure containing chlorine atoms alternately in left- and right-hand positions.[3,4] The parts of the molecules in which chlorine atoms are indiscriminately placed are probably in the amorphous regions, and this of course is consistent with the properties of chlorinated polythene and with the idea that chlorine and hydrogen atoms are too different in size to be indiscriminately interchanged in a crystal structure. By contrast, polytrifluorochloroethylene $(-CF_2-CFCl-)_p$ does crystallize quite well; it is not known whether there is any stereochemical regularity such as that which occurs to some extent in polyvinyl chloride, but even if there is not, there is much less difference in size between fluorine and chlorine atoms (radii 1.35, 1.70A) than between hydrogen and chlorine (radii 1.1, 1.70A), so that even stereochemically irregular molecules would probably be able to crystallize.

Hitherto, attention has been focused on regularity of molecular structure as a criterion for crystallization. But is it the only criterion? Is it possible for molecules of completely regular structure to be incapable of

[1] C. W. Bunn, Nature **161**, 929 (1948).

[2] C. W. Bunn and H. S. Peiser, Nature **161**, 159 (1947).

[3] R. Brill and F. Halle, Naturwiss. **26**, 12 (1938).

[4] C. S. Fuller, Chem. Revs. **26**, 143 (1940).

UNIVERSITY OF MICHIGAN LIBRARIES

crystallizing? It is sometimes suggested that molecules of very awkward shape may pack so badly that crystallization will not occur. Certainly, in monomeric organic crystals Kitaigorodski[5] found that the molecular packing fraction (the fraction of the space occupied by the molecules, assuming that each atom has a spherical boundary of definite radius) varies considerably from one substance to another; and he suggested that the failure to crystallize of some alkyl naphthalene derivatives (which harden to glasses instead of crystallizing) might be due to the poor packing capabilities found in some members of this class of substances. The matter deserves some consideration: is it a reasonable suggestion at all, and if so, what types of molecules are likely to pack badly?

In the first place, the approach to the problem of crystallizability by way of the concept of economical packing is only valid for molecules which do not form localized strong bonds; when hydrogen bonds are formed, quite open crystal structures may be stable; the extreme example is ice, the formation of which from water is accompanied by a 10 percent expansion, in contrast to the 10–15 percent contraction which occurs when most nonpolar liquids crystallize. It is the lower energy content of the crystal which is the significant fact; however, for nonpolar molecules the energy content always decreases with the molecular volume, and therefore for these substances it is valid to consider crystallizability in terms of molecular packing efficiency.

In the second place, it is not the molecular packing fraction in the crystal alone which is important, but the difference between the packing fractions in the crystal and the liquid. The crucial question is this: is it possible for the most close-packed regular arrangement of any molecular species to be less dense than the irregular arrangement in the liquid? It seems unlikely, though not necessarily impossible. If it *is* possible for any species, then that molecular species will not crystallize.

The problem of the relation of packing efficiency to shape is a difficult one. Certainly packing efficiency has nothing to do with symmetry: a jigsaw puzzle composed of completely unsymmetrical pieces packs perfectly, whereas one can imagine very symmetrical shapes which would pack very badly. Nor is it any simple function of general over-all shape such as the relative dimensions in different directions. In chain polymers, one might expect to find the lowest packing efficiencies in substances whose molecules have long side groups which are not closely crowded along the chains—molecules with a skeletal form; but the possibilities would have to be examined individually—some skeletal molecules might pack very well, if the side groups fit in well with those of neighboring molecules, while others may inevitably leave gaps. However, my own impression is that bad packing is an unlikely cause of failure to crystallize, for the reason already given—that it depends on the difference of the packing efficiencies of irregular and regular arrangements and it has yet to be demonstrated that the packing efficiency in the regular arrangement can be the lower for any shape of molecule. It is much more likely that when a regular polymer is reported as noncrystalline, the situation is not that it is incapable of crystallizing, but that it does so only with difficulty, and the right conditions have not been found. This brings us to the next topic—the molecular characteristics which determine how readily a polymer crystallizes.

Crystallization involves two different processes—nucleus formation, and subsequent growth of the nuclei; it is with the first that we are primarily concerned, though both enter into the problem. Although no quantitative work has been done, it appears from general experience that in chain polymers the probability of nucleus formation is related to temperature in the same way as in monomeric substances: as the temperature falls below the melting point, the probability of nucleus formation rises to a maximum and then falls again, ultimately reaching zero. (Fig. 1 a). If a specimen can be cooled sufficiently rapidly through the nucleus-forming temperature range, it will set to a glass, and at a sufficiently low temperature can be kept for very long periods without crystallizing. If the temperature of the glasslike specimen is raised gradually, a point will be reached at which crystallization starts. The crystallization temperature appears to be related to the "relaxation transition" or "second-order transition" temperature; this is the point at which, on raising the temperature, the specific heat and coefficient of expansion of a glasslike polymer suddenly increase and the specimen becomes more flexible and somewhat rubberlike; there appears to be a fairly sudden increase in molecular mobility at this temperature. However, the degree of molecular mobility just above the relaxation transition temperature is not sufficient for the formation of crystal nuclei: in three aromatic polyesters Kolb and Izard[6] found that crystallization occurs at an appreciable rate at temperatures 30 to 50°C above the respective relaxation transition temperatures, and Reinhardt[7] records that in polyvinylidene chloride, for which the relaxation

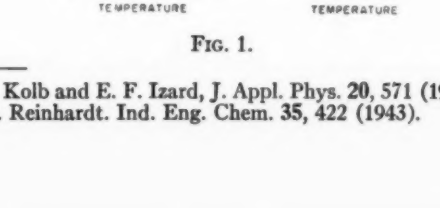


FIG. 1.

[5] A. Kitaigorodski. Acta Physicochimica URSS 22, 309 (1947).

[6] H. J. Kolb and E. F. Izard, J. Appl. Phys. 20, 571 (1949).

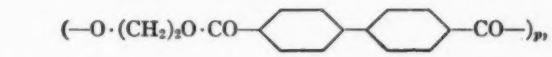
[7] R. C. Reinhardt. Ind. Eng. Chem. 35, 422 (1943).

transition temperature is −17°C, crystallization even at room temperature is very slow.

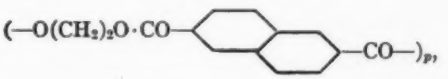
The general situation thus appears to be of the form sketched in Fig. 1 a: nucleus formation occurs only over a limited temperature range *SN* which is narrower than the interval *RM* between relaxation transition and melting temperatures. The crystallization range *SN*, and the magnitude of the probability of nucleus formation (height of the maximum) within that range, depend on the interval *RM* and on the intervals *RS* and *NM*; anything that reduces the interval between relaxation transition and melting temperatures reduces the crystallization range, and anything that increases the intervals *RS* and *NM* likewise reduces the crystallization range and may well decrease the maximum probability *P* of nucleus formation attained within that range (Fig. 1 b). If the range *RM* is short enough, and the intervals *RS* and *NM* occupy a considerable part of it, crystal nucleus formation may have a low probability at any temperature, so that lengthy annealing might be necessary at a temperature which is difficult to find. It is therefore necessary to ask, in the first place, what molecular characteristics control relaxation transition temperatures and melting points; and, in the second place, the intervals *RS* and *NM*.

The molecular characteristics which determine relaxation transition temperatures are not known with any certainty; nor are those which control melting points. Both these subjects present large and difficult problems. On the relation between these temperatures, it has been observed empirically by Boyer[8] and by Beaman[9] that the ratio of relaxation transition to melting temperatures (absolute) is low (~0.5) for polythene and other polymers containing two identical substituents on the same chain atom, but has a higher value (0.6 to 0.7) for polymers with less symmetrical molecules (such as gutta-percha and rubber). There is, however, no consistent correlation with molecular symmetry: there are exceptions in both directions, for polyethylene terephthalate, a highly symmetrical molecule, shows the ratio 0.64, and polytrifluorochloroethylene, which has lower symmetry, shows the ratio 0.53.

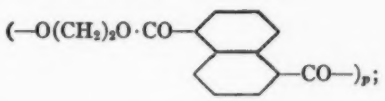
It will probably be more readily possible to understand the relaxation transition temperatures of amorphous substances in terms of molecular characteristics than to understand the melting points of crystalline substances. Both these temperatures are likely to be controlled by the magnitude of the forces between the molecules and by the flexibilities of the mlecules (as measured by the energy necessary to deform them), but the melting points of crystals are likely to depend also on highly specific factors such as the efficiency of propagation of vibrations through the crystal structures. A survey of the melting points of substances consisting of rigid monomeric molecules indicates that general overall shape has a considerable influence: *meta* di-derivatives of benzene and *cis*-isomers of ethylene derivatives melt lower than *para*-di-derivatives and *trans*-ethylene derivatives, respectively, and there are many examples to show that it is not a matter of symmetry in the strict sense, but of general overall shape; moreover, molecules of skeletal shape melt lower than comparable molecules of compact shape. We do not know how these shape factors operate; but whatever the mechanism, we may expect that similar factors influence polymer melting points, for in a general way polymer melting points conform with those of comparable monomeric substances. (A detailed discussion of these points will be published elsewhere.) This expectation seems to be realized in the few polymer examples available: natural rubber, the *cis*-isomer of polyisoprene, melts lower than the *trans*-isomer gutta-percha, and among polyesters and polyurethanes containing CH=CH groups, Marvel and Young[10] found that the *cis*-isomers melt lower than the *trans*. The influence of "streamlined" shape is perhaps shown by comparing the ethylene glycol polyesters of *p,p'*-diphenylene dicarboxylic acid



2,6 napthalene dicarboxylic acid

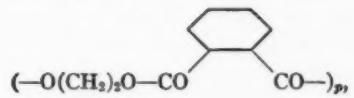


and the corresponding 2,5 compound

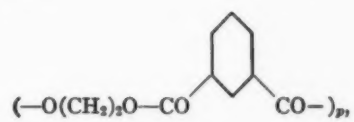


the first melts at 346°C, the second at 260°C and the third at 230°C.[11] The relaxation transition temperatures of such comparable substances usually differ less than do the melting points; thus, for the less streamlined molecules, the interval between relaxation transition and melting temperatures is the smaller.

The ethylene glycol polyesters of phthalic acid



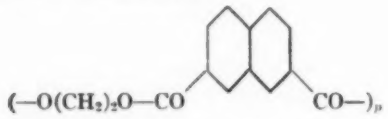
isophthalic acid



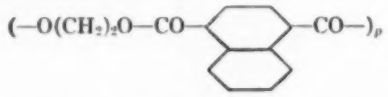
[8] R. F. Boyer. J. Appl. Phys. **25**, 825 (1954).
[9] R. G. Beaman. J. Polymer Sci. **9**, 472 (1952).
[10] C. S. Marvel and C. H. Young. J. Am. Chem. Soc. **73**, 1066 (1951).
[11] R. Hill and E. E. Walker. J. Polymer Sci. **3**, 609 (1948).

UNIVERSITY OF MICHIGAN LIBRARIES

2,7 naphthalene dicarboxylic acid

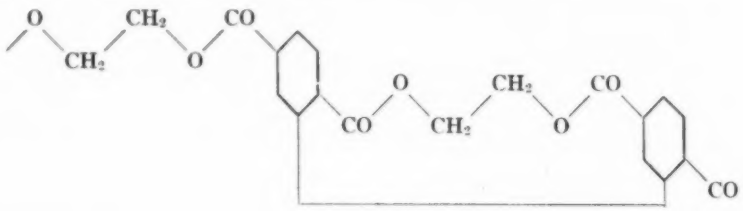


and 1,4 naphthalene dicarboxylic acid

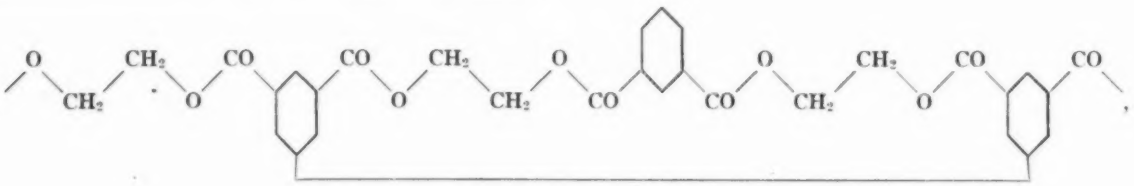


have not been obtained in crystalline form.[11] From what has just been said, one would expect the melting points of these polymers (assuming that they are capable of crystallizing at all) to be considerably lower than those of the more symmetrical analogs, and the intervals between relaxation transition and melting points to be correspondigly narrowed. The crystallization range would therefore be restricted, and this may be part of the reason why these substances have not been crystallize.

The crystallization range, however, depends also on the intervals *RS* and *NM* of Fig. 1. The width of the interval *RS*—the difference of temperature between the relaxation transition temperature *R* and the temperature *S* at which crystallization occurs—is a matter of molecular mobility and of the chance of a molecule getting into the right position for forming a crystalline arrangement with its neighbors. For a given degree of molecular mobility (which is attained a certain number of degrees above the relaxation transition temperature), the chance of a molecule getting into the right position with respect to its neighbors will depend on the length of the chain repeat distance. Simple molecules like polythene and polytetrafluoroethylene, if they are not in the appropriate position for crystallization, have not far to go to get there, and this may be the reason why it is very difficult to get them in amorphous form. Polyethylene terephthalate, with its longer repeat distance (10.75 A), can be obtained in amorphous form by rapid cooling, though when cooled slowly it crystallizes well. An important point of difference between the terephthalate and isophthalate polymers is that, whereas the chain repeating unit of the former consists of only one chemical unit



the repeating unit of the latter (assuming that it keeps to a more or less flat configuration in the crystal, as in the terephthalate crystals) consists of two chemical units



and therefore, for a given degree of molecular mobility, molecules have much less chance of getting into the right positions for crystallization. The interval *RS* may therefore be greater than for the terephthalate, and the crystallization range correspondingly reduced. Moreover, the rate of growth of crystals may also be much reduced for the same reason. The same considerations apply to the other polymers mentioned above (such as 2,7 and 1,4 naphthalates); it may be that for all these polymers the crystallization range is much reduced, the probability of nucleus formation low at any temperature (as in Fig. 1 b), and the rate of growth also low; and these may be the reasons why these polymers have not been obtained in crystalline form.

A particularly interesting substance in the consideration of this subject is polyisobutene

$$(-CH_2-C(CH_3)_2-)_p,$$

a rubberlike substance at room temperature which, in the unstretched state, has never been crystallized by cooling. There is here no question of inability to crystallize, for on stretching it crystallizes remarkably well, giving a sharp and highly detailed x-ray diffraction pattern. The relaxation transition temperature is given at $-74°C$.[12] Since the substance has not been crystallized by cooling, the melting point is unknown; it must be below room temperature, however, and probably below 0°C. A good deal of work has been done

[12] R. F. Boyer and R. S. Spencer. *Advances in Colloid Science* (Interscience Publishers, Inc., New York 1946), Volume II, p. 1.

on the properties of the substance below 0°C, and if it had crystallized, this must have been noticed. The ratio of relaxation transition temperature to melting point is evidently unusually high, and the interval between the two unusually small; this may well be a case of the type of Fig. 1 b. The chain repeat distance in the crystal is long (18.6 A) and comprises 8 chemical units, but in this case it is doubtful whether the great length of the repeating unit is an important contributing factor, for the chemical chain units are very short and the configuration is probably some sort of helix, so that if a molecule is not suitably placed for crystallization, only minor changes of configuration (such as occur in the usual way by random thermal movements) are necessary to make it so; large translations along the chain axis are not necessary. Admittedly, if the molecular configuration in the crystal is complex, the chance of attaining that configuration by random thermal movements is less than if the configuration is simple; but changes of configuration can occur progressively along the chain, and such a process is inherently more likely than a large bodily translation of a molecule. These considerations suggest that nucleus formation and crystal growth depend on both the length of the chemical chain unit and the length of the geometrical repeating unit.

To sum up: for chain polymer molecules to be able to crystallize at all, they must have a geometrically regular structure or an approximation to it. It is doubtful whether any substance of quite regular molecular structure is incapable of crystallizing, but some regular polymers crystallize only with difficulty, and this is associated with a small interval between relaxation transition and melting points. The temperature range in which crystallization proceeds at an appreciable rate may be still further restricted if the chain repeat distance of the molecules is long; and for the same reason, crystallization at any temperature may be slow.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 7 JULY, 1954

Relationship of First- to Second-Order Transition Temperatures for Crystalline High Polymers*

RAYMOND F. BOYER
The Dow Chemical Company, Midland, Michigan
(Received January 12, 1954)

This paper concerns an empirical linear relationship between the melting temperature T_1 and the second-order or glassy transition temperature T_2 for partially crystalline polymers. Two lines can be drawn, one for symmetrical polymers, the other for unsymmetrical polymers, the latter being of lower slope. These families are only approximate. It is shown for a limited group of polymers, that T_2 is proportional to the cohesive energy density of the polymer. Since $T_1=\Delta H/\Delta S$, it is suggested that the two families of curves arise because of differences in ΔS which depends on symmetry and chain stiffness. It is shown that T_1 and T_2 are linearly related with diluent as a common parameter.

INTRODUCTION

WHEREAS all amorphous polymers appear to have a second-order transition,[1-3] crystalline polymers claim the additional interest of possessing both a first- and a second-order transition.[4,5] This paper is concerned with the facts and implications of an empirical linear relationship which we have discovered connecting these two transition temperatures, T_1 and T_2, respectively, of crystalline high polymers.

We reported informally on these results at the Columbus meeting of the High Polymer Division of the American Physical Society (March 20, 1952), while a detailed report was presented at the 2nd International Conference on Physical Chemistry, Paris, June 6, 1952. This material was available in preprint form for this meeting and was subsequently published in book form.[6] Beaman[7] has discovered this same general relationship independently of us. Jenckel[7a] has likewise commented on this same general relation, although his findings are essentially a restatement of what Kauzmann had reported for inorganic glasses.[8]

We define these two transition temperatures by means of Fig. 1, which is a schematic volume-temperature curve closely patterned after the one by Bekkedahl for Hevea rubber.[5] The transition occurring at the higher temperature arises from the melting of the

* Presented at the Symposium on Crystallization of Polymers at the Division of High-Polymer Physics, American Physical Society Meeting, Durham, North Carolina, March 27, 1953.

[1] E. Jenckel, Kolloid-Z. **100**, 163 (1942).
[2] R. F. Boyer and R. S. Spencer, *Advances in Colloid Science* (Interscience Publishers, Inc., New York, 1946), Vol. 2, p. 1 ff.
[3] E. Jenckel, Kolloid-Z. **120**, 174 (1951).
[4] M. Ruhemann and F. Simon, Z. physik. Chem. **A138**, 1 (1938).
[5] N. Bekkedahl, J. Research Natl. Bur. Standards **13**, 411 (1934).
[6] *Changements de Phases* (Societe de Chimie Physique, Paris, 1952), p. 383 ff.
[7] R. G. Beaman, J. Polymer Sci. **9**, 472 (1953).
[7a] E. Jenckel, Kolloid-Z. **130**, 64 (1953).
[8] W. Kauzmann, Chem. Revs. **43**, 219 (1948).

UNIVERSITY OF MICHIGAN LIBRARIES

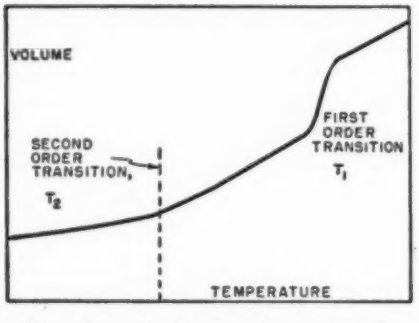


FIG. 1. Schematic volume-temperature curve for a partially crystalline polymer, used to define the first- and second-order transition temperature, T_1 and T_2.

rubber crystallites, whereas T_2 represents a change on heating from a glassy solid to a highly viscous liquid.

The melting point, while a true thermodynamic transition, is not precisely defined, but depends on the previous thermal history of the sample, and especially on the temperature used in freezing the rubber. This has been shown by Bekkedahl and Wood[9] who also found a broad range in melting point for any given temperature of freezing. Freezing and melting are both time-dependent, and it is generally not possible to freeze the polymer completely. The three points mentioned in this paragraph have been described in great detail by Wood[10] and are a consequence of the ong chain nature of the polymer. Recent unpublishedl work of Roberts and Mandelkern[11] has shown that there is an equilibrium T_1 for rubber. However, this does not invalidate the statement that many published melting points have been subject to uncertainty. The melting point increases with molecular weight but approaches a limiting value at higher molecular weights. We shall be concerned here primarily with that asymptotic value.

The second-order transition temperature is not a true thermodynamic transition: there is no latent heat associated with it and apparently only a relatively simple change in x-ray pattern.[12] Its value depends on the speed of the test used to observe it, being higher the faster the speed of the test. For this reason it has been referred to as the "apparent second-order transition," and the "glassy transition." A critical examination of the theoretical aspects of these phenomena has been presented by Kauffmann,[8] while Boyer and Spencer, and Jenckel[1-3] have reviewed much of the experimental data in the field.

The second-order transition is apparently confined to the amorphous areas of a crystalline polymer and the value of T_2 is only slightly influenced by the presence of crystallinity.[5] The magnitude of the change in slope of the volume temperature curve at T_2 is less the greater the amount of crystallinity. For highly crystalline polymers the presence of the break may be very difficult to observe. In general, the magnitude of this break will be a measure of the amount of crystallinity present in the sample.

Since the second-order transition point is rate-dependent, it is necessary to choose some standard time. Most volume-temperature curves are run at a reasonably fixed rate of heating—usually around 1°C per minute. In some cases we have been forced to employ dielectric data, but have tried to extrapolate from the frequencies used back to a frequency of about 10^{-2} cps, following Kauffmann.[8] Such extrapolations are doubtful because the activation energy is changing so rapidly as the transition is approached.[13,14] and different polymers differ in this regard.[15]

The second-order transition temperature will likewise increase with increasing molecular weight M, as first shown by Ueberreiter, reaching a limiting value as indicated by the following expression of Fox and Flory:[13]

$$T_2 = a + b/M, \qquad (2)$$

where a and b are constants. Again, one should use the asymptotic value of T_2 but cannot be certain that values in the literature are so. The first-order transition temperature is the ratio $T_1 = \Delta H/\Delta S$, where ΔH, the heat of fusion, depends largely on intermolecular forces, while ΔS, the entropy of fusion, probably depends mainly on geometrical considerations.[16]

There is no general theory linking T_2 with chemical composition primarily because the theory of the liquid state has not been worked out as yet. Pauling, by treating

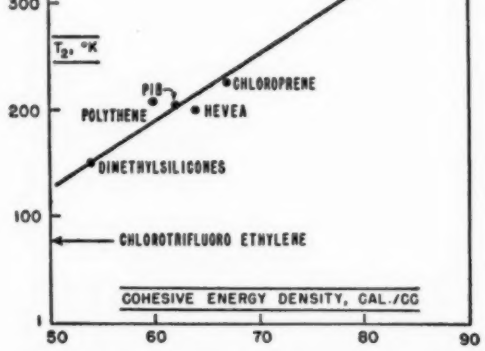


FIG. 2. Dependence of second-order transition temperature on the cohesive energy density of the polymer.

[9] N. Bekkedahl and L. A. Wood, J. Chem. Phys. **9**, 193 (1941).
[10] L. A. Wood, *Advances in Colloid Science* (Interscience Publishers, Inc., New York, 1946), Vol. 2, p. 57 ff.
[11] D. Roberts and L. Mandelkern, Durham, North Carolina, Meeting, American Physical Society, March 1953.
[12] S. Krimm and A. V. Tobolsky, Textile Research J. **21**, 805 (1951).
[13] T. G. Fox, Jr. and P. J. Flory, J. Appl. Phys. **21**, 581 (1950).
[14] Bischoff, Catsiff, and Tobolsky, J. Am. Chem. Soc. **A74**, 3378 (1952).
[15] See reference 6, Fig. 10, p. 398.
[16] H. Mark, J. Appl. Phys. **12**, 41 (1941).

second-order transitions in crystalline solids as rotational transitions, showed that high intermolecular forces or high moment of inertia would lead to a high transition temperature.[17] While a direct application of Pauling's criteria to glassy polymers is probably not justified, it at least affords a point of departure. Figure 2 shows a plot of T_2 as a function of cohesive energy density for a series of noncrystalline polymers. In general, it would appear the intermolecular forces are of prime importance in governing the magnitude of T_2, although other factors such as geometry and hindered rotation around C−C bonds must play some role.[1,18]

Figure 3 shows the relation between T_2 and T_1, for all polymers on which we were able to find measured values. In general, it appears that the data follow two well-defined patterns—the symmetrical polymers lying along the upper line, and the unsymmetrical polymers following along the lower line. The lines, as drawn, have slopes of 2.75 and 1.50. It will be noted that some of the condensation polymers fall on the upper line, others on the lower line. The value of T_1 for polyisobutylene was on a stretched sample.

We have stated the source of these various T_1 and T_2 values in a previous publication.[6] Omitted from that tabulation, but included on the present chart, were the values for sebacoyl piperazine determined by Flory, Mandelkern, and Hall.[19] $T_1=180°C$, $T_2=82°C$. In addition, we have changed T_2 for Nylon from 45°C, based on dielectric measurements, to a value of −20°C reported by Morgan.[20] Several new values will be reported in the following discussions.

It is almost inevitable that each point on Fig. 3 will be subject to revision when definitive values of T_1 and T_2 are available. In the meanwhile, debate about the validity of each pair of T_1, T_2, values is to be expected. Nowhere is this better illustrated than by Teflon. The T_2 value of 20°C is generally described as a first-order transition because of its reversible change in x-ray pattern.[20a] At the same time a glasslike transition for this polymer has been reported at about 160°K.[21] This latter point would be completely off the curve. To complicate this picture still further, Sack[22] has reported that the change in mechanical properties of Teflon at T_2 is characteristic of a glassy transition. Bunn has suggested[23] that the room temperature transition in Teflon may actually represent a rotational

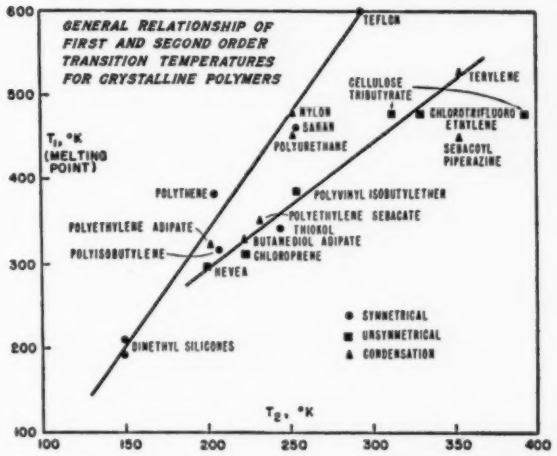


FIG. 3. Relationship between melting point T_1 and second-order transition temperature T_2. The upper line covers the symmetrical poymenrs, the lower ones the unsymmetrical polymers. The coldensation polymers show divided allegiance.

transition in the crystal lattice involving entire polymer molecules. This rotation would be possible according to Bunn, because of the tubelike nature of the molecule and the large hindrance to rotation about single bonds. One might then tentatively conclude that this rotation on the crystalline lattice is primarily responsible for the location of the high melting temperature of Teflon and that the point as indicated should fall on the curve. Bunn stated[23] that the intermolecular forces in Teflon are quite low, which would be consistent with the T_2 value of 160°K.

Chlorotrifluoroethylene likewise presents an anomalous overall picture. Hoffman had suggested[24] that T_2 might be as low as −80°C, and indeed this would be consistent with the extremely low cohesive energy density of this polymer, namely 50 cal/cc as reported by Bueche.[25] This is the lowest known cohesive energy density for any polymer and, if plotted on an extension of Fig. 2, would place T_2 in the range of −140°C. It may well be that the fluorinated polymers would constitute a new class of materials lying on a line of their own, because of the absence of hydrogen bonding forces. The actual value plotted on Fig. 3 is 47°C, which was determined by Redding.[26] Redding noted that on highly crystalline samples he could not obtain a well-defined break. We had obtained a value of −20°C from a single volume-temperature curve.[6]

The pattern exhibited by Fig. 3 might be explained broadly on the basis that T_1 and T_2 are both proportional to ΔH, and that the several families of curves arise because T_1 also varies as $1/\Delta S$.

Alfrey has proposed[27] that part of this difference

[17] L. Pauling, Phys. Rev. **36**, 430 (1930).
[18] See reference 6, Tables I through IV.
[19] Flory, Mandelkern, and Hall, J. Am. Chem. Soc. **73**, 2535 (1951).
[20] Dr. L. B. Morgan, Imperial Chemical Industries, Ltd., Manchester, England (private communication).
[20a] H. A. Rigby and C. W. Bunn, Nature **164**, 583 (1949); Quinn, Roberts, and Work, J. Appl. Phys. **22**, 1085 (1951).
[21] Furakawa, McCoskey, and King, J. Research Natl. Bur. Standards **49**, 273 (1952).
[22] Prof. H. Sack, Cornell University (private communication).
[23] C. W. Bunn, Durham, North Carolina Meeting, American Physical Society, March 1953.
[24] J. D. Hoffman, Abstract R-2., Columbus, Ohio Meeting, American Physical Society, March 1952.
[25] A. M. Bueche, J. Am. Chem. Soc. **74**, 65 (1952).
[26] Dr. F. P. Redding, Carbide & Carbon Chemicals Corp. (private communication).
[27] Dr. T. Alfrey, Jr., The Dow Chemical Company, Midland, Michigan.

UNIVERSITY OF MICHIGAN LIBRARIES

may result because the unsymmetrical polymers possess the mixed *d-l* configurations in the sense proposed by Huggins,[28] whereas the symmetrical polymers cannot. Thus the unsymmetrical ones act like copolymers with larger ΔS and lower melting point.

Baker and Smyth[29] have shown (for simple organic molecules) that symmetrical compounds which can rotate in the solid state and thereby gain rotational entropy have a much higher melting point than unsymmetrical compounds of similar chemical composition. We have questioned[6] whether this same effect is acting for the symmetrical polymers of Fig. 3. To do so, rotation in the crystal lattice or some coupling between the amorphous areas where the second-order transition has already occurred and the crystalline region where the first-order transition is about to occur, would be needed.

It is possible, as mentioned earlier, that the room temperature transition in Teflon does represent rotation of polymer segments or molecules in the crystal lattice and is responsible for the very high melting point of this polymer. We have attempted to examine dielectric data in a search for evidence of premelting anomalies. Curiously enough, the only dielectric data we could find going through the melting temperature is that of Wurstlin on a poly 1, 4-butanediol adipate.[30] His results might be interpreted either as rotational effects in the crystalline lattice prior to melting, or to a broad distribution of melting points.

The other question of interplay between amorphous and crystalline areas can arise because the same polymer chains pass from amorphous into crystalline areas with a direct mechanical linkage. The specific question would then be: If the amorphous areas of a symmetrical polymer have gained more rotational entropy at a given temperature than the amorphous areas of an unsymmetrical polymer, will the former have a higher melting point?

The answers to many of the questions posed by this correlation will probably have to await more precise values of T_1 and T_2, as well as experimental values of the heats and entropies of fusion, and cohesive energy densities for more polymers.

RELATIVE DEPENDENCE OF T_1 AND T_2 ON DILUENT

Addition of a diluent to a polymer will depress both the first- and the second-order transition temperatures, and the question arises (first posed to us by Dr. L. A. Wood) as to whether the linear relation between T_1 and T_2 would still be maintained, with diluent content as the common parameter. According to Flory,[31] the melting temperature T_1 will decrease with diluent according to

$$1/T_1-1/T_1^0=k_1v_1(1-BV_1v_1/RT_1), \qquad (1)$$

where T_1^0 is the melting point of the pure polymer. k_1 is a constant containing the heat of fusion of the polymer crystallites, B is the energy of mixing in cal/cc between polymer and diluent, while V_1 is the molar volume of the diluent which is present in volume fraction, v_1. If B is zero, the right-hand side of the equation varies linearly with volume fraction of diluent, or

$$1/T_1-1/T_1^0=k_1v_1. \qquad (2)$$

If $B\neq 0$, then, as they have shown,

$$(1/T_1-1/T_1^0)1/v_1\sim v_1/T_1. \qquad (3)$$

The effect of diluents on T_2 is more obscure. Fox and Flory[13] suggested that T_2 depends on the number of free chain ends, whence T_2 might be expected to decrease linearly with volume fraction of diluent. Fox[32] has elaborated further on this point, as have Mandelkern and Flory for the special case of crystalline polymers.[33]

We have reviewed[34] some of these questions and have suggested that since the second-order transition is a rate phenomenon, it might well represent a state for which the diffusion constant of the diluent in the polymer is constant. In analogy with Eqs. (15) and (17) of our paper, we can write

$$1/T_2-1/T_2^0=k_2v_1/E_d, \qquad (4)$$

where k_2 is a constant and E_d is the activation energy for diffusion.

If the polymer-solvent interaction constant B is zero, it follows from a comparison of Eq. (2) and (3)

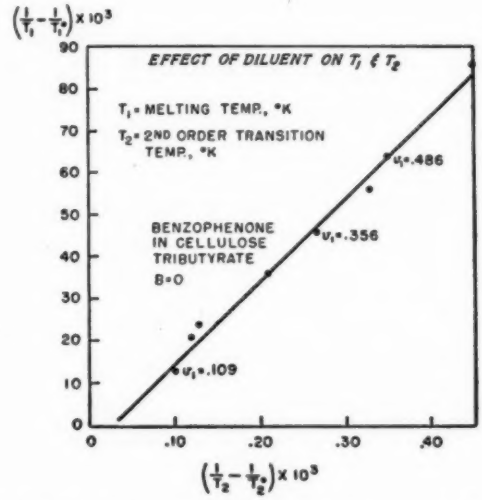


FIG. 4. Relation between T_1 and T_2 in the presence of diluent. plotted according to Eq. (5). v_1 is the volume fraction of diluent,

[28] M. L. Huggins, J. Am. Chem. Soc. **66**, 1991 (1939).
[29] W. O. Baker and C. P. Smyth, J. Am. Chem. Soc. **61**, 1695 (1939).
[30] F. Wurstlin, Kolloid- Z. **120**, 84 (1951), especially Fig. 5.
[31] P. J. Flory, J. Chem. Phys. **17**, 223 (1949).
[32] T. G. Fox, Jr., Bull. Am. Phys. Soc. **27**, 19 (1952).
[33] L. Mandelkern and P. J. Flory, J. Am. Chem. Soc. **73**, 3210 (1951).
[34] R. F. Boyer, J. Appl. Phys. **22**, 723 (1951).

that

$$(1/T_2-1/T_2^0)k_3/E_d=1/T_1-1/T_1^0, \qquad (5)$$

where k_3 is a new constant. This representation for benzophenone in cellulose tributyrate is shown in Fig. 4 based on the data of Mandelkern and Flory. The upper of the two values of T_2 was used. If $B\neq 0$, then one should plot

$$(1/T_1-1/T_1^0)1/v_1\sim 1/T_2-1/T_2^0 \qquad (6)$$

which neglects the variation in T_1 on the right side of the equation. This does indeed give a linear plot for ethyl laurate, but there are only five points on the curve. Mandelkern[35] has pointed out that the percent of crystallinity can change with amount of diluent present so that Eqs. (5) and (6), unless v_1 is corrected for the amount of amorphous material present, cannot be given a rigorous interpretation. The simple plots of t_1 against t_2 for three solvents of different B value are shown in Fig. 5, again using the data of Mandelkern and Flory.[33]

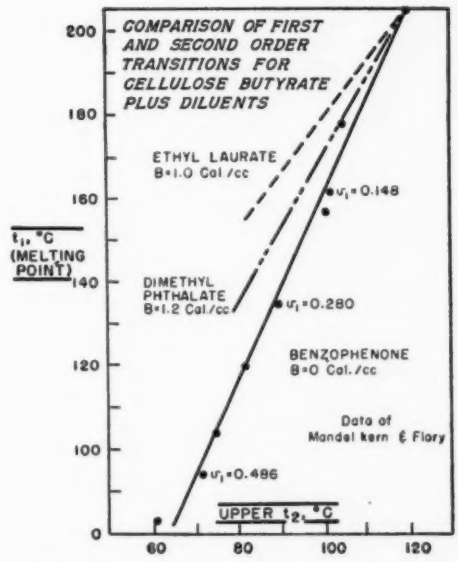


FIG. 5. Relationship between melting point and second-order transition temperature in °C for cellulose tributyrate containing diluents. v_1 is the volume fraction of diluent, B is the energy of interaction between polymer and diluent.

Other Parameters

Since both T_1 and T_2 increase with molecular weight in about the same manner, a linear relation between them for a series of fractions of different molecular weight might be anticipated. To our knowledge there is no data available for testing this point. Also, both T_1 and T_2 increase with applied pressure, and some correlation should be expected, although this type of data is also lacking.

ACKNOWLEDGMENTS

The author would like to acknowledge several interesting discussions with Dr. Turner Alfrey and Dr. Robert Simha on the subject matter of this paper. We are also indebted to Dr. L. A. Wood for raising the question of the effect of diluents.

[35] L. Mandelkern, National Bureau of Standards (private communication).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 7 JULY, 1954

Crystallization Kinetics in High Polymers. I. Bulk Polymers*†

LEO MANDELKERN AND FRED A. QUINN, JR., *National Bureau of Standards, Washington 25, D. C.*
AND
PAUL J. FLORY, *Department of Chemistry, Cornell University, Ithaca, New York*

(Received January 12, 1954)

The rates of crystallization of five polymers, poly-(ethylene oxide), poly-(decamethylene adipate), and three poly-(N,N′-sebacoyl piperazine)'s differing in molecular weight, have been observed dilatometrically at various temperatures ranging from 5.5 to 24° below the respective melting points T_m. The crystallization isotherms are remarkably reproducible, and the previous thermal history above the melting point is without effect provided that melting is completed before the crystallization experiment is begun. The crystallization isotherms display induction periods. The very large temperature coefficients observed are uniform throughout the crystallization process, i.e., isotherms at different temperatures are superimposable by rescaling the time. It is inferred from this and other evidence that steady-state nucleation and growth occur simultaneously throughout the process. The results are satisfactorily interpreted by applying conventional nucleation theory and assuming that nuclei grow at rates proportional to their surface areas. The treatment resembles that applied to crystallization in metals. Spherical and disk-shaped nuclei are considered, but the data do not discriminate between them. An extension of the theory accounts for the maximum commonly observed in the crystallization rate as the temperature is decreased below the range of the present measurements.

INTRODUCTION

PREVIOUS investigations[1–6] of melting in crystalline polymers have demonstrated the existence of an equilibrium melting temperature T_m, and it has been shown that this temperature is depressed in a systematic manner by diluents, copolymerized units, and chain ends. The transition of crystalline to amorphous polymer may be treated by thermodynamic methods appropriate to two phase equilibrium, although the rate of attainment of true equilibrium may be extremely slow. The polycrystalline morphology in a given instance obviously must depend on the conditions prevailing during crystallization. Here kinetic factors rather than thermodynamic equilibrium considerations are of primary importance. Several experimental studies of crystallization rates in polymers have been reported,[7–10] and it is apparent from these investigations that the rate is strongly dependent on the temperature.

The experiments reported in the present paper were undertaken after it was discovered, in conjunction with melting and annealing studies, that highly reproducible crystallization rates could be realized. Since an understanding of the nature of the nucleation process seemed an obvious first objective, the experiments were limited to the temperature range just below the melting point, wherein nucleation plays a dominant role. Several different polymers have been employed in order to gain an indication of the generality of the conclusions reached. Volume dilatometers were used to study the development of crystallinity, inasmuch as they are easily maintained at carefully controlled temperatures for the required long periods of time and the volume variations are very sensitive measures of changes in the crystallinity.

EXPERIMENTAL

Materials

Poly-(ethylene oxide), "PEO," was obtained from the Carbide and Carbon Chemicals Company; its number average molecular weight, as determined by end group analyses, was in the range 9000–10 000. After refluxing with ethyl ether for 24 hours to remove the relatively small amounts of low molecular weight material present, it was dried *in vacuo* at 40°C. The density of the amorphous polymer at 100°C, as determined by means of a pipette pycnometer,[11] was 1.0699 ± 0.0005 g/cm³. Its intrinsic viscosity at 30°C in carbon tetrachloride was 0.23 dl/g.

Poly-(decamethylene adipate), "PDA," was prepared[6] by the polymerization of carefully purified decamethylene glycol and adipic acid at 160° to 217°C under a reduced pressure of nitrogen for 33 hours. The viscosity at 109°C was 1800 poises, corresponding to a number average molecular weight of 11 500.[6] Its density at 109°C as determined by the pipette pycnometer was 0.9732 ± 0.0004 g/cm³.

The polyamide, poly-(N,N′-sebacoyl piperazine), "pip-10," was prepared in a manner similar to that

* This work was supported in part by the U. S. Office of Naval Research and the Office of the Quartermaster General.

† Presented at the Symposium on Crystallization of Polymers at the Division of High-Polymer Physics, American Physical Society Meeting, Durham, North Carolina, March 27, 1953.

[1] P. J. Flory, J. Chem. Phys. **17**, 223 (1949).
[2] Evans, Mighton, and Flory, J. Am. Chem. Soc. **72**, 2018 (1950).
[3] Flory, Mandelkern, and Hall, J. Am. Chem. Soc. **73**, 2532 (1951).
[4] L. Mandelkern and P. J. Flory, J. Am. Chem. Soc. **73**, 3206 (1951).
[5] A. M. Bueche, J. Am. Chem. Soc. **74**, 65 (1952).
[6] Mandelkern, Garrett, and Flory, J. Am. Chem. Soc. **74**, 3949 (1952).
[7] L. A. Wood and N. Bekkedahl, J. Appl. Phys. **17**, 362 (1946).
[8] F. P. Price, J. Am. Chem. Soc. **74**, 311 (1952).
[9] W. H. Cobbs, Jr., and R. L. Burton, J. Polymer Sci. **10**, 275 (1953).
[10] P. W. Allen, Trans. Faraday Soc. **48**, 1178 (1952).
[11] P. J. Flory, J. Am. Chem. Soc. **62**, 1057 (1940).

TABLE I. Molecular parameters governing the temperature dependence of the crystallization rate.

Polymer	T_m °C	ΔH_u cal/mole of unit	$\Delta h_u \times 10^{-9}$ ergs/cm³ of unit	Slope	$\sigma_S^2\sigma_E$ ergs³/cm⁴	σ_S[a] ergs/cm	$\sigma_S/\Delta H_u$[b]
PEO	66	1980	2.08	−4.44	238	7.8	0.13
PDA	79.5	10 200	1.51	−3.40	98	5.8	0.13
Pip-10A	173	6200	0.957	−7.28	86	5.6	0.19
Pip-10B	182	6200	0.957	−7.28	86	5.6	0.19
Pip-10C	183	6200	0.957	−6.16	73	5.3	0.14

[a] For $\sigma_E = 0.5\sigma_S$.
[b] For surface thickness of 3A and $\sigma_E = 0.5\sigma_S$.

previously described.[3] Three samples, differing widely in their melt viscosities, were prepared.[12] Sample *A* had a melt viscosity of about 10 poises at 200°C; sample *B*, whose properties have been described previously,[3] had a melt viscosity of 800 poises at 190°C, corresponding to a number average molecular weight of about 10 000. Sample *C* had a melt viscosity of 2000 poises at 250°C. The specific volume of sample *B* at 189°C has been reported as 0.9263±0.0007 cm³/g, and this value has been assumed for the other two pip-10 samples. Since the relative volume changes, and not their absolute values, are important, this assumption is inconsequential in the present work.

Dilatometry

The specific volumes were measured in mercury-filled dilatometers using techniques previously described.[3,4] A constant-temperature oil bath regulated to within ±0.1°C was used. A modified thyratron relay[3] assured constant operation of the bath for the long periods of time required for the work reported here.

For studies of the kinetics of the crystallization an auxiliary oil bath controlled to within ±0.5°C was used in conjunction with the thyratron-regulated thermostat. The dilatometers were initially held in the auxiliary bath at a specified temperature above the melting temperature, sufficient time being allowed for all crystallites to melt. They were then quickly transferred to the more accurately controlled thermostat which was set at a predetermined temperature below the melting temperature, and readings were taken at specified time intervals while the temperature was maintained constant.

RESULTS

Melting Temperature and Heats of Fusion

The specific volumes of the bulk polymers were measured in the order of ascending temperature, and T_m was determined as the temperature at which the last traces of crystallinity disappeared. The previous thermal histories of the samples consisted only in the gradual uncontrolled cooling of the melt at the time of preparation. The specific volume-temperature plots for PDA and pip-10B have already been published.[3,6] The temperature-time sequences followed for pip-10A and pip-10C were identical to that previously described for pip-10B, and the same behavior was observed. Volume equilibrium for PEO was attained in 1 to 2 hours at all temperatures except in the range less than 10° below T_m. In this interval the temperature was raised in 0.5° or 1.0° increments and held from 12 to 24 hours at each temperature. Volume equilibrium appeared to be established within such periods. In PEO, the melting took place over a comparatively narrow temperature range; 90 percent of the volume change occurred in a 5° interval, a behavior similar to that recently reported for polymethylene.[13]

The heat of fusion per mole of repeating unit, ΔH_u, was determined by studying the depression of T_m by diluents at various concentrations according to the methods previously described and demonstrated.[1,3] The results for PDA and pip-10 have already been published and a detailed account of the determination of ΔH_u for PEO will appear in a forthcoming paper.[14] The values of T_m and ΔH_u thus determined are given in columns 2 and 3 of Table I.

Crystallization Kinetics

The isothermal rates of crystallization of the supercooled polymeric liquids were investigated within an

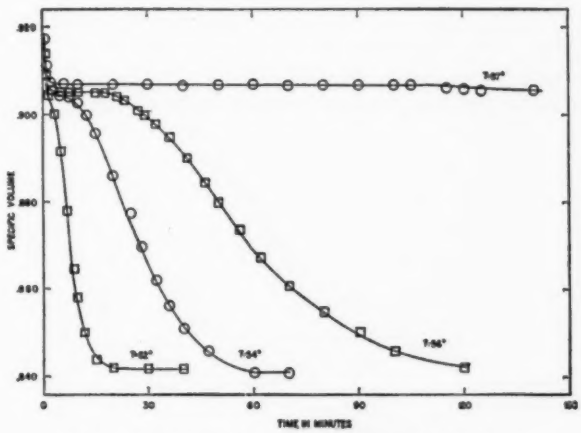


FIG. 1. Specific volume *vs* time for poly-(ethylene oxide) at various temperatures.

[12] We are indebted to Dr. Helen P. Bedon for the preparation of samples *A* and *C*.

[13] Mandelkern, Hellman, Brown, Roberts, and Quinn, J. Am. Chem. Soc. **75**, 4093 (1953).
[14] L. Mandelkern (to be published).

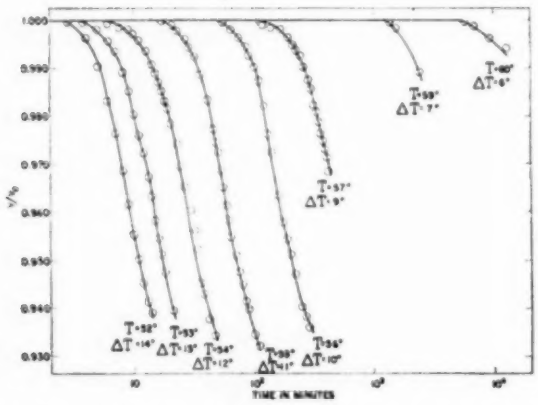


FIG. 2. Ratio of specific volume at time t to that at zero time, plotted against log time for poly-(ethylene oxide) at various temperatures.

appropriate interval somewhat below T_m. The exact interval was chosen so that the times τ_i at which crystallization became detectable were in the range from 5 to 5×10^3 min. The dilatometer dimensions were such that volume changes of the order of 0.02 percent could be detected, corresponding to a degree of crystallinity of about 0.2 percent for a polymer displaying a 10 percent change in volume on melting.

Typical results of the dependence of the specific volume of PEO on time are given in Fig. 1 for various temperatures in the range of interest. Curves of this type were also obtained for PDA and for all three pip-10's. Similar isothermal crystallization curves have been observed by Wood and Bekkedahl[7] for natural rubber and by Allen[10] for poly-(hexamethylene adipamide). Beyond the initial period of temperature equilibration the volume remains constant for a well-defined "apparent induction time" τ_i. Crystallization becomes observable at τ_i, as indicated by the decrease in volume, and the process proceeds at an accelerating rate immediately thereafter. Ultimately the volume attains a constant value, as a pseudoequilibrium degree of crystallization is approached. These general observations of the crystallization process are similar to those

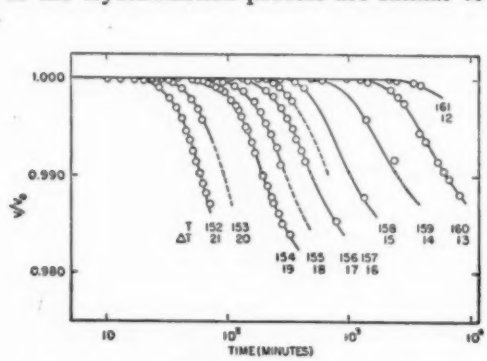


FIG. 3. Ratio of specific volume at time t to that at zero time, plotted against log time for poly-(N,N'-sebacoyl piperazine)-A at various temperatures.

discussed by Wood,[15] and are also similar to the results for polyethylene terephthalate[9] reported by Cobbs and Burton. The latter investigators used infrared techniques to measure the development of crystallinity.

The experiments reported here are highly reproducible. Almost identical volume-time curves were obtained by successively repeating an experiment with a given sample. As will be shown later, duplicate samples of the same polyamide give nearly identical rate constants. The results are also independent of the temperature at which the sample is initially held above T_m, provided this temperature is maintained for a sufficient period of time to insure the complete disappearance of crystallinity. Thus, for a range of one to fifty degrees above T_m no superheating effect was found in contrast to the results reported for poly-(hexamethylene adipamide)[10] and poly-(chlorotrifluoroethylene).[8]

The very strong effect of temperature on the crystallization rate is illustrated in Fig. 1, where τ_i varies by a factor of 15 in the interval from 54° to 57°C. In Table II are given the values of τ_i for various ΔT's, representing the number of degrees below the melting temperature at which the observations were made. The marked dependence of τ_i on temperature in this region near T_m is again apparent. If $\log\tau_i$ is plotted against $\log\Delta T$ a family of straight lines is obtained with slopes which indicate that the empirical relation $\tau_i=\text{const}(1/\Delta T)^9$ approximates the observations. By extrapolating this relationship to a ΔT of one degree, the time necessary for the onset of crystallization to be observed becomes the order of 10^6 years.

TABLE II. Apparent induction time τ_i at various values of ΔT (the number of degrees below the melting temperature).

PEO		PDA		Pip-10A		Pip-10B		Pip-10C	
ΔT °C	τ_i min	ΔT °C	τ_i min	ΔT °C	τ_i min	ΔT °C	τ_i min	ΔT °C	τ_i min
6	6000	5.5	6900	12	2100	12	4200	12	1100
7	1200	6.5	850	13	950	13	2000	13	700
8	320	7.5	255	14	450	14	1065	14	420
9	110	8.5	84	15	245	15	700	15	250
10	55	9.5	35	16	130	16	288	16	145
11	21	10.5	11	17	100	17	170	17	87
12	7			18	56	18	103	18	65
13	4			19	38	19	80	19	40
				20	24	20	62	20	28
						21	40	21	17
						22	35	22	12
						23	20		
						24	12		

The entire crystallization process is illustrated in Figs. 2, 3, 4, and 5 for PEO and for the three pip-10's, respectively. The ratio V/V_0 of the volume at time t to the initial volume at zero time is plotted in these figures against $\log t$ at the various temperatures of

[15] L. A. Wood, *Advances in Colloid Sciences* (Interscience Publishers, Inc., New York, 1946), Vol. II, p. 63 ff.

interest. The previously described characteristics of isothermal crystallization are maintained in these plots; the magnitude of the "apparent induction times" can be seen along the upper axis representing $V/V_0=1$; the acceleration of the crystallization and the leveling off are apparent. PDA behaves in a similar manner, but at most temperatures the crystallization was allowed to proceed only slightly beyond τ_i.

DISCUSSION

Stable nuclei exceeding a definite critical size must be formed by means of statistical thermal fluctuations[16–18] in order for crystallization to occur in a homogeneous, supercooled liquid. Nuclei of the critical size are formed through growth of smaller, subcritical nuclei.[19–22] If the sample is brought suddenly to the temperature $T(<T_m)$ at which crystallization is subsequently to proceed, the population density of these subcritical nuclei initially will not immediately reach the steady state level. At the very outset subcritical nuclei approaching the critical size will be virtually nonexistent, and the initial rate of generation of stable nuclei must, therefore, be zero. Only after an incubation time τ^*, during which subcritical nuclei are developing in the quantity allowed by thermodynamic equilibrium, will the generation of stable nuclei commence. Subsequently, a steady state rate will be reached.

The volume-time plots shown in Figs. 1 to 5 would seem to suggest, as has previously been postulated for the case of poly-(hexamethylene adipamide),[10] that the time τ_i at which crystallization is first observed may be identified with τ^*. It is evident, of course, that the observed time τ_i will depend on the sensitivity of the crystallinity detector; if dilatometers less sensitive to volume changes were used, τ_i would occur at later times at all temperatures, while τ^* would remain constant. The change would, of course, be small, inasmuch as the acceleration of the rate is great in this region. Although for condensed systems it is difficult to make a theoretical estimate of the magnitude of τ^*, it can be shown that it must vary inversely as $(\Delta T)^4$ for spherical[21] or for disk-shaped nuclei. The observed dependence on $(\Delta T)^{-9}$ obviously cannot be reconciled with this interpretation. On this basis alone the identification of τ_i with τ^* would seem unlikely, and this indication is supported by evidence to be cited below.

An important property of the crystallization process is illustrated in Figs. 2, 3, 4, and 5. For a given polymer the volume-log time plots at all temperatures can be

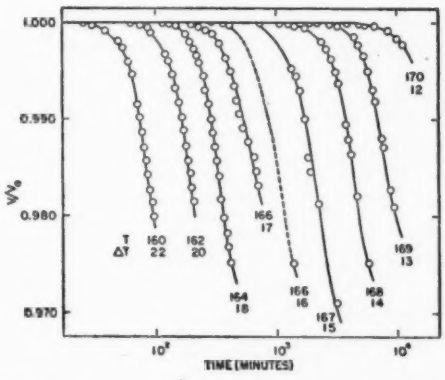


FIG. 4. Ratio of specific volume at time t to that at zero time, plotted against log time for poly-(N,N′-sebacoyl piperazine)-B at various temperatures.

made to coincide merely by shifting each of the curves an appropriate distance along the horizontal axis. The superposition of the isotherms holds for the entire crystallization process from zero time to completion. The similar superposition of crystallization isotherms holds also[23] for gutta percha, polymethylene, and polyethylene, and therefore appears to be general for the crystallization of polymers in the temperature range near the melting point. Thus the temperature dependence of the crystallization process remains invariant with the extent of crystallization. Evidently, by the application of suitable time-scale factors the volume-time data taken at different temperatures can be made to coincide. The superposability of the isotherms and the consequent constancy of the temperature coefficient throughout the course of the crystallization weighs heavily against any mechanism consisting of separate incubation and growth phases. Concurrent nucleation and growth receives strong support from this observation. We postulate, therefore, that τ^* constitutes but a negligible part of the observed time τ_i, i.e., that the steady state rate of nucleation is quickly attained.

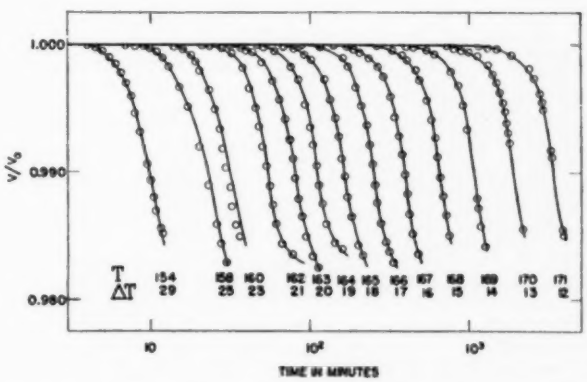


FIG. 5. Ratio of specific volume at time t to that at zero time, plotted against log time for poly-(N,N′-sebacoyl piperazine)-C at various temperatures.

[16] R. Becker and W. Döring, Ann. phys. [5] **24**, 719 (1935).
[17] D. Turnbull and J. C. Fisher, J. Chem. Phys. **17**, 71 (1949).
[18] R. S. Bradley, Quart. Revs. (London) **5**, 315 (1951).
[19] D. Turnbull, Metals Technol., Am. Inst. Mining Met. Engrs. Tech. Publ. No. 2365, 1948.
[20] R. F. Probstein, J. Chem. Phys. **19**, 619 (1951).
[21] A. Kantrowitz, J. Chem. Phys. **19**, 1097 (1951).
[22] J. R. Zeldovitch, Acta Physicochim. U.R.S.S. **18**, 1 (1943).
[23] Mandelkern, Quinn, and Roberts, unpublished data.

A homogeneous nucleation process is indicated by the excellent reproducibility of duplicate observations made either on the same sample or using different samples of the same polymer; it is supported also by the absence of an effect of superheating temperature on the subsequent crystallization rate. Homogeneous nucleation implies that stable nuclei are formed in the bulk of the supercooled liquid; foreign bodies, cavities, or other heterogeneities of spurious origin which can act as centers of nucleation presumably make no appreciable contribution to the observed crystallization kinetics. Once formed, a stable nucleus must grow. Lacking any direct evidence as to the detailed nature of the growth mechanism, it is further assumed that the radius of a growing center increases linearly with time, and hence that the rate of volume change of a growing center is proportional to its surface area.

These postulates satisfy the general requirements demanded by the superposition of the isotherms; they are identical to those proposed by Johnson and Mehl[24a,b] in their treatment of the crystallization kinetics of metals.

The rate of generation of nuclei and the highly significant variation of this rate with temperature may depend on the shape of the critical nucleus. In the following treatment we shall adopt two alternative models, the one consisting of a spherical nucleus and the other disk-shaped. The latter may be assumed to comprise a bundle of sections of polymer chains oriented parallel to the axis of the disk.

The characteristic shape of the isotherm describing the conversion of amorphous to crystalline polymer will depend also on the form of development of the nucleus during its subsequent growth. Spherical nuclei will be assumed to give rise to centers which continue to develop spherically. Disk-shaped nuclei will be assumed either to grow along their radii, with the polymer chain normal to the plane of the disk, or to develop spherically, i.e., to grow in all directions equally. The occurrence of spherical growth (irrespective of the initial form of the nucleus) is supported by the common observations of spherulites[8,25–28] in crystalline polymers. Each of these spherulites consists of numerous smaller crystallites (and intervening amorphous regions) which radiate from a common center, the direction of crystal growth being along the radius of the spherulite.[25] The center may be presumed to mark the site of generation of a nucleus. Studies of the magnitude and sign of the birefringence of these structures indicate that the long axes of the molecules are perpendicular to the radii of the spherulites. Growth can then be envisaged as occurring by means of the lateral aggregation of segments of the chain molecules.

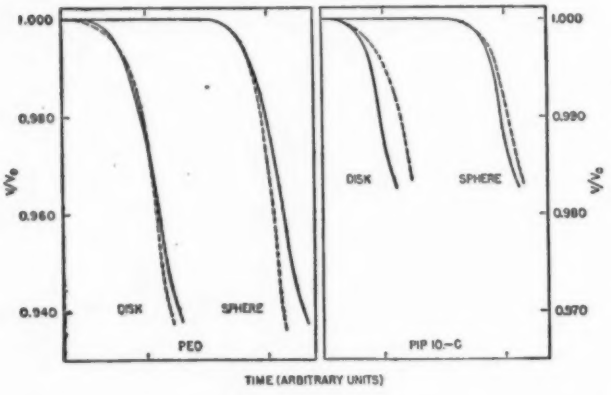


FIG. 6. Fit of experimental data to theoretical curves. Ratio of specific volume at time t to that at zero time, plotted against log time for poly-(ethylene oxide) and poly-(N,N′-sebacoyl piperazine)-C. Solid lines experimental results; dotted lines appropriate theoretical curves. Position of each isotherm relative to abscissa is arbitrary; short vertical marks along abscissa indicates length of one decade.

A major consideration in the calculation of the rate at which crystallinity develops is the effect of the impingement of growing centers upon each other.[24a,b;29a,b] Impingement may be taken into account conveniently according to the method of Avrami by the following device. We first assume a hypothetical analog in which the nuclei are allowed to form and grow in the transformed as well as in the untransformed mass. The nuclei that are allowed to form in the transformed mass are termed "phantom nuclei" since, of course, they actually do not exist. The actual crystallization rate is then recalculated by considering only that portion of the increase of mass of an average growing center which takes place in space not already transformed.

Let dW_c' be the mass that is transformed during time dt, in the approximation that there is no impingement of growing centers. The actual mass transformed dW_c will be less than dW_c' because of the impingement. The shells of transformed material are in random regions since nuclei, including the phantoms, have been formed at random throughout the entire mass. Therefore, in the vicinity of the growing centers the fraction of the mass that is untransformed is the same as the total fraction untransformed. Then the expected actual mass transformed is given by

$$dW_c = (W_l/W_0)dW_c', \tag{1}$$

where W_l/W_0 is the fraction of the mass untransformed at time t. W_0 is the total mass of the system and W_l is the mass untransformed at time t. A more rigorous mathematical derivation of (1) is given by Avrami.[29a,b]

[24a] W. A. Johnson and R. F. Mehl, Trans. Am. Inst. Mining Met. Engrs. **135**, 416 (1939).
[24b] W. A. Johnson and R. F. Mehl, Metals Technol., Am. Inst. Mining Met. Engrs. Tech. Publ. No. 1089, August, 1939.
[25] C. W. Bunn and T. C. Alcock, Trans. Faraday Soc. **41**, 317 (1945).
[26] W. M. D. Bryant, J. Polymer Sci. **2**, 547 (1947).
[27] S. W. Hawkins and R. B. Richards, J. Polymer Sci. **4**, 515 (1949).
[28] A. Keller, Nature **169**, 913 (1952); ibid. **171**, 170 (1953).
[29a] M. Avrami, J. Chem. Phys. **7**, 1103 (1939).
[29b] M. Avrami, J. Chem. Phys. **8**, 212 (1940).

TABLE III. Crystallization rate constants for various polymers at various values of ΔT (the number of degrees below the melting temperature).

Poly-(ethylene oxide) $\delta=0.917$			Poly-(decamethylene adipate) $\delta=0.917$			Poly-(N,N'-sebacoyl piperazine)-A $\delta=0.964$			Poly-(N,N'-sebacoyl piperazine)-B $\delta=0.964$			Poly-(N,N'-sebacoyl piperazine)-C $\delta=0.964$		
ΔT °C	k_s	k_d	ΔT °C	k_s	k_d	ΔT °C	k_s	k_d	ΔT °C	k_s	k_d	ΔT °C	k_s	k_d
6	6.3×10^{-18}	6.4×10^{-14}	5.5	5.1×10^{-18}	3.3×10^{-14}	12	6×10^{-16}	1.1×10^{-12}	12	1.4×10^{-17}	1.1×10^{-13}	12	3.8×10^{-15}	8.0×10^{-12}
7	1.4×10^{-14}	8.1×10^{-12}	6.5	1.4×10^{-14}	1.7×10^{-11}	13	3.4×10^{-15}	6.2×10^{-12}	13	2.9×10^{-16}	1.0×10^{-12}	13	3.4×10^{-14}	3.9×10^{-11}
9	2.7×10^{-11}	7.0×10^{-9}	7.5	6.2×10^{-13}	2.8×10^{-10}	14	9.9×10^{-14}	8.7×10^{-11}	14	3.9×10^{-15}	8.6×10^{-12}	14	3.5×10^{-13}	1.5×10^{-10}
10	1.2×10^{-9}	1.3×10^{-7}	8.5	5.8×10^{-11}	1.1×10^{-8}	15	1.6×10^{-12}	6.0×10^{-10}	15	6.7×10^{-14}	4.9×10^{-11}	15	2.8×10^{-12}	1.0×10^{-9}
11	1.0×10^{-7}	2.9×10^{-6}	9.5	2.7×10^{-9}	1.9×10^{-7}	16	1.3×10^{-11}	2.9×10^{-9}	16	3.7×10^{-13}	2.7×10^{-10}	16	1.9×10^{-11}	4.6×10^{-9}
12	3.6×10^{-6}	1.4×10^{-5}	10.5	7.7×10^{-9}	1.5×10^{-6}	17	3.9×10^{-11}	8.0×10^{-9}	17	1.2×10^{-11}	2.5×10^{-9}	17	1.0×10^{-10}	1.5×10^{-8}
13	3.9×10^{-5}	2.6×10^{-4}				18	2.3×10^{-10}	3.0×10^{-8}	18	9.0×10^{-11}	1.4×10^{-8}	18	5.4×10^{-10}	5.5×10^{-8}
						19	3.8×10^{-9}	1.0×10^{-7}	20	1.1×10^{-9}	7.6×10^{-8}	19	2.5×10^{-9}	1.8×10^{-7}
						21	1.1×10^{-7}	2.9×10^{-6}	22	2.5×10^{-8}	7.9×10^{-7}	20	1.0×10^{-8}	4.9×10^{-7}
												21	4.9×10^{-9}	1.2×10^{-6}
												23	5.8×10^{-8}	2.9×10^{-6}
												25	2.9×10^{-6}	2.5×10^{-5}

We postulate that the radius of a growing center increases linearly with time. Then, in the approximation of no impingement, the volume $v_c'(t,z)$ at time t of a given spherically growing center (irrespective of whether the nucleus from which it was formed is spherical or disk-shaped) which was initiated at time z is

$$v_c'(t,z)\cong(4\pi/3)G^3(t-z)^3, \tag{2}$$

where G, which we shall term the growth rate constant, is the rate of change of the radius with time. For a disk of thickness l_c growing only along the radius

$$v_c'(t,z)\cong\pi l_c G^2(t-z)^2. \tag{3}$$

The mass W_c', transformed at time t without regard to impingement (but allowing phantom nuclei to form and grow), is given by

$$\begin{aligned} W_c'(t)&=A'W_0\int_0^t w_c'(t,z)dz \\ &=W_0A(\rho_c/\rho_l)\int_0^t v_c(t,z)dz, \end{aligned} \tag{4}$$

where $w_c'(t,z)$ is the mass at time t of a given center initiated at time z, A' is the rate of formation of nuclei per unit of mass, and A is the corresponding rate per unit of volume. The density of the liquid polymer is ρ_l and that of the totally crystalline polymer is ρ_c.

By substituting Eq. (2) or (3) into (4), and the resultant into (1), we obtain by integration

$$\ln\{W_0/(W_0-W_c)\}=(\pi/3)AG^3(\rho_c/\rho_l)t^4, \tag{5}$$

for spherical growth, and

$$\ln\{W_0/(W_0-W_c)\}=(\pi/3)l_cAG^2(\rho_c/\rho_l)t^3 \tag{6}$$

for disk growth.

Since the volume is measured experimentally, more convenient forms of Eqs. (5) and (6) are obtained by making the following substitutions:

$$\begin{aligned} W_0&=\rho_l V_0, \\ W_c&=\rho_c V_c, \\ V_c&=[V-V_0][\delta/(\delta-1)], \end{aligned}$$

where V is the total volume at time t, V_0 is the initial volume of the supercooled liquid, V_c is the volume transformed at time t and $\delta=\rho_l/\rho_c$. Then

$$\ln\{(\delta-1)/[\delta-(V/V_0)]\}=(\pi/3\delta)AG^3t^4=k_st^4 \tag{7}$$

$$\ln\{(\delta-1)/[\delta-(V/V_0)]\}=(\pi/3\delta)l_cAG^2t^3=k_dt^3. \tag{8}$$

The dependence of V/V_0 on time and temperature, as described by Eqs. (7) and (8), is in accord with the observed superposability of the experimental isotherms (Figs. 2, 3, 4, and 5). The mechanisms assumed thus satisfy the necessary condition of superposition. If there is no latent volume effect on crystallization, then of course $\delta=1$, and the analysis leads to equations which are identical to those derived by Johnson and Mehl.[24a,b] For the case where crystallinity develops from a fixed number of heterogeneities Avrami[29a,b] has deduced equations which are similar in form to (7) and (8). However, in his case the physical significance of k_s and k_d is different; while the shape of the growing centers still governs the time dependence of the isothermal crystallization, the exponent of the time in his equations is not a constant, but varies as crystallinity develops.

The expected variation of (V/V_0) with time can be calculated from Eqs. (7) and (8) for a given value of δ, (δ being typical of the polymer), and comparison may then be made with the experimental data. The values of δ, given in Table III for the polymers under consideration, were estimated from the equilibrium specific volume-temperature curves and the extrapolated liquid lines, assuming that at a temperature far removed from T_m the equilibrium degree of crystallinity was unity. Small errors in the estimated magnitude of $\delta$and in its variation with temperature have a negligible effect on the subsequent analysis.

A comparison of the theory with the experimental data is presented in Fig. 6 for PEO and pip-10C. The solid lines represent the average curves obtained by superposition of the experimental isotherms of Figs. 2 and 5; the dotted lines are the appropriate theoretical curves, the calculation of which involves the use of only one arbitrary parameter, k_s or k_d, which determines the position of the isotherms relative to the log$t$ axis.

UNIVERSITY OF MICHIGAN LIBRARIES

The degree of correlation shown is typical of the various crystallization isotherms studied. Quantitative agreement is obtained for the initial stages of the process for either spherical or disk-type growth; the experimental data are not sufficiently precise to distinguish between them. At higher degrees of crystallization the agreement between theory and experiment is only qualitative. The crystallization rate constants k_s and k_d, given in Table III, decrease enormously with temperature.

Polymeric systems in general do not crystallize completely even at equilibrium. As the crystallization progresses, therefore, not all the liquid or amorphous region can be transformed. In calculating the rate at which mass is transformed, this effect must be considered along with the effect of the impingement of growing centers.

Let dW_c' be the increase, in time dt, of the transformed mass, neglecting both the effect of impingement and the fact that crystallization will not go to completion. As crystallinity develops, newly formed crystallites will be unable to grow in portions of the amorphous region, just as they are unable to grow in regions actually transformed. Then, by analogy with Eq. (1) dW_c, the actual increase in mass transformed, becomes

$$dW_c = dW_c'[1-U(t)]. \tag{1'}$$

$U(t)$ is "the effective fraction" of the mass transformed at time t; it is that fraction of the total mass, at time t, in which further crystal growth is forbidden. The effective fraction transformed is assumed to be proportional to the actual fraction transformed, the proportionality factor being $(1/\lambda_w)$, the reciprocal of the equilibrium degree of crystallinity computed on a mass basis. Taking

$$U(t) = (W_c/W_0)(1/\lambda_w), \tag{9}$$

it is found, employing methods identical with the case previously discussed, that

$$\ln\{\lambda_w(\delta-1)/[\lambda_w(\delta-1)-((V/V_0)-1)]\} = (1/\lambda_w)k_s t^4 \tag{7'}$$

$$\ln\{\lambda_w(\delta-1)/[\lambda_w(\delta-1)-((V/V_0)-1)]\} = (1/\lambda_w)k_d t^3. \tag{8'}$$

Equations (7') and (8') are similar in form to Eqs. (7) and (8), and they reduce to them exactly when λ_w is unity. In Fig. 7, V/V_0 is plotted against $\log(k_s^{\frac{1}{4}}t)$ and against $\log(k_d^{\frac{1}{3}}t)$ for various values of λ_w for δ equal to 0.917. The effect of $\lambda_w<1$ is minor except for low values of λ_w, where the shapes of the (V/V_0)-log time plots are altered appreciably as a result of the large proportion of the amorphous regions in which crystal growth is ultimately forbidden.

Since the polymers investigated were all sharp-melting, there was no appreciable changes in λ_w with temperature in the interval of interest. The physical properties of the polymers are indicative of relatively high values of λ_w; hence errors in the absolute values of k_s or k_d resulting from the use of Eqs. (7) and (8) were insignificant. Though in the systems studied here the corrections for incomplete crystallization are small, these corrections could be important in analyzing the crystallization kinetics in other systems. For example, in copolymers or in polymer-diluent mixtures λ_w may be small; also the melting range will be considerably broadened, hence λ_w may change even over a narrow temperature interval.

The temperature coefficient of the rate constant is governed by the temperature coefficients of both the growth rate and the steady state nucleation rate. The free energy charge required for the formation of a small nucleus within the bulk of a mother liquor is the resultant of the contributions of two terms: the bulk free energy which is negative and the surface free energy which is positive.[16-18] The free energy, ΔF_s, of formation of a spherical nucleus of radius r is given by

$$\Delta F_s = 4\pi r^2\sigma_S - (4\pi/3)r^3\Delta f_0, \tag{10}$$

where σ_S is the interfacial free energy per unit area between crystal nucleus and liquid, and Δf_0 is the bulk free energy of fusion per unit volume. For the formation of a disk of thickness l and radius r

$$\Delta F_d = 2\pi rl\sigma_S + 2\pi r^2\sigma_E - \pi r^2 l\Delta f_0, \tag{11}$$

where σ_S and σ_E are the interfacial free energies per unit area for the cylindrical surface and the two end surfaces, respectively.

In the classical development of nucleation theory the free energy of formation possesses a maximum with respect to nucleus size. The maximum value of the free energy of formation, ΔF^*, is the height of the free

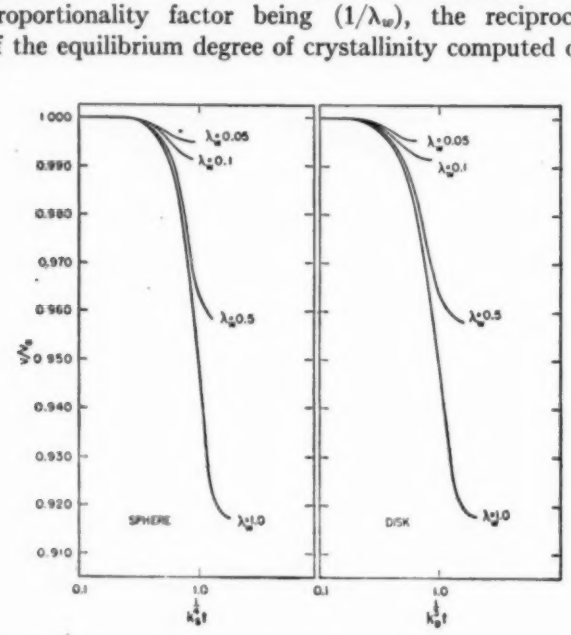


FIG. 7. Theoretical plot of ratio of specific volume at time t to that at zero time, plotted against log $(k_s^{\frac{1}{4}}t)$ and log $(k_d^{\frac{1}{3}}t)$ for various values of λ_w.

energy barrier that must be surmounted in order for stable nuclei to be formed. Nuclei which are larger than the size corresponding to ΔF^* can grow spontaneously, while smaller nuclei are thermodynamically unstable and disappear. Maximizing Eq. (10) with respect to r and equating Δf_0 to $(\Delta T/T_m)\Delta h_u$, where Δh_u is the heat of fusion per unit volume of repeating unit, there is obtained

$$\Delta F_s^* = (16\pi/3)[\sigma_S^3/(\Delta h_u^2)]T_m^2/(\Delta T)^2. \quad (12)$$

For disk-shaped nuclei ΔF_d^* is the value at the saddle point of the free energy surface described by Eq. (11). The coordinates of this point are obtained by equating both $(\partial \Delta F_d/\partial r)_l$ and $(\partial \Delta F_d/\partial l)_r$ to zero and solving the resultant simultaneous equations for r and l.[30] At the saddle point

$$\Delta^* F_d = [8\pi\sigma_S^2\sigma_E/(\Delta h_u^2)]T_m^2/(\Delta T)^2. \quad (13)$$

The statistical equilibrium theory[1] of crystallization of segments of polymer molecules may also be used to calculate ΔF^*. The free energy of forming a semicrystalline array from the supercooled polymeric liquid has been calculated in this theory.[31] For the case where there is no diluent present and the thickness of the nucleus is appreciably less than the length of the completely extended molecule, the value calculated in this way for ΔF^* is of the same form as that given in Eq. (13).

Becker and Döring[16] and Turnbull and Fisher[17] have shown that the rate of nucleation in condensed systems may be expressed by

$$A = A_0 \exp\{-E_D/\mathbf{k}T - \Delta F^*/\mathbf{k}T\}, \quad (14)$$

where $\mathbf{k}$ is Boltzmann's constant, A_0 is a constant which is only slightly temperature dependent, and E_D, the apparent activation energy of viscous flow per molecule, serves as an approximation for the energy of activation for transport across the liquid-crystal interface. Assuming the growth to be mainly diffusion controlled, then

$$G = G_0 \exp\{-E_D/\mathbf{k}T\}. \quad (15)$$

For polymeric liquids E_D changes with temperature over any extended range.[32,33] Both E_D and G_0 can be assumed to be temperature independent over the small temperature intervals of present interest.

For spherical nucleus formation and spherical growth it is found from Eqs. (7), (12), (14), and (15) that

$$k_s^s = (\pi/3\delta)A_0G_0^3 \times \exp\{-4E_D/\mathbf{k}T - (16\pi/3) \times [\sigma_S^3/(\Delta h_u)^2]T_m^2/(\Delta T)^2\mathbf{k}T\}. \quad (16)$$

[30] R. Courant, *Differential and Integral Calculus* (Nordemann Publishing Company, New York) Vol. II, p. 206.
[31] Equation (19), reference 1.
[32] T. G. Fox, Jr., and P. J. Flory, J. Appl. Phys. **21**, 581 (1950).
[33] T. G. Fox, Jr., and P. J. Flory, J. Phys. and Colloid Chem. **55**, 221 (1951).

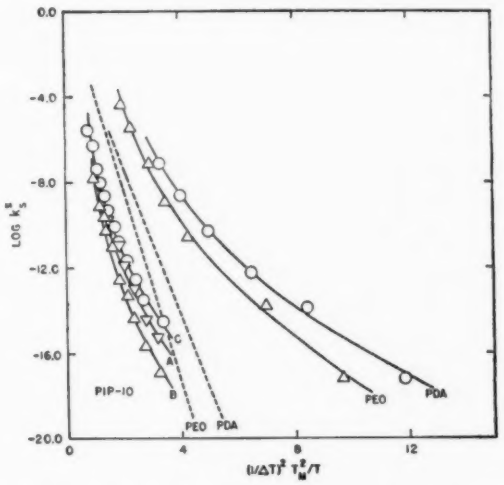


FIG. 8. Plot of log k_s^s against temperature variable $(1/\Delta T)(T_m^2/T)$ for five polymers. Dashed lines represent results using values of T_m three degrees higher than those listed in Table I.

For the formation of disk nuclei accompanied by disk-type growth we obtain from Eqs. (8), (13), (14), and (15)

$$k_d^d = (\pi l_c/3\delta)A_0G_0^2 \exp\{-3E_D/\mathbf{k}T - [8\pi\sigma_S^2\sigma_E/(\Delta h_u)^2]T_m^2/(\Delta T)^2\mathbf{k}T\}. \quad (17)$$

Similarly, for the formation of disk nuclei accompanied by spherical growth

$$k_s^d = (\pi/3\delta)A_0G_0^3 \exp\{-4E_D/\mathbf{k}T - [8\pi\sigma_S^2\sigma_E/(\Delta h_u)^2]T_m^2/(\Delta T)^2\mathbf{k}T\}. \quad (18)$$

An equation similar in form to these has been deduced by Ookawa[34] from other considerations. As is apparent from the derivation of Eqs. (7) and (8), the form of the isotherm depends only on the nature of the growth process. Hence the values deduced for either k_s^s or k_s^d for a given crystallization experiment will be identical. Aside from minor differences in the first term in the exponential in the expression for k_d^d, all three mechanisms lead to the same temperature coefficient. Over the small temperature range of the present experiments the pre-exponential factor may be considered to remain constant and we may write

$$\log k_s^s = B - 7.28\sigma_S^3T_m^2/(\Delta h_u)^2(\Delta T)^2\mathbf{k}T, \quad (16')$$

where

$$B = \log[(\pi/3\delta)A_0G_0^3] - 4E_D/2.3\mathbf{k}T.$$

The other rate constants may be similarly expressed; differences occur only in the values for B, in the numerical factors in the second terms of the equations corresponding to Eq. (16'), and in the replacement of σ_S^3 by $\sigma_S^2\sigma_E$.

Plots of $\log k_s^s$ (or $\log k_s^d$) and $\log k_d^d$ against $T_m^2/T(\Delta T)^2$ are shown in Figs. 8 and 9, respectively. In both plots the results for the three pip-10's could be approximated by straight lines, but those for PEO

[34] A. Ookawa, J. Phys. Soc. Japan **6**, 473 (1951).

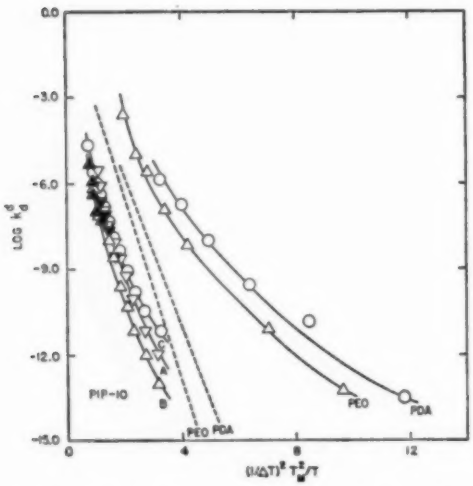


FIG. 9. Plot of log k_d^d against temperature variable $(1/\Delta T)^2 (T_m^2/T)$ for five polymers. Duplicate sample of poly-(N,N'-sebacoyl piperazine)-B▲. Duplicate sample of poly-(N,N'-sebacoyl piperazine)-A▼. Dashed lines represent results using values of T_m three degrees higher than those listed in Table I.

and PDA definitely are curved. However, the abscissas in these plots are extremely sensitive to small errors in the values of ΔT. An uncertainty remains in the correct absolute value for T_m. The temperature required is that temperature at which perfect crystallites would melt at equilibrium. It has been assumed that this temperature may be identified with the value obtained by the slow heating of a sample previously crystallized by ordinary (i.e., rapid) cooling. Recent experimental results[35,36] indicate, however, that under the most carefully controlled condition of crystallization the true melting temperature may be as much as three or four degrees higher than values heretofore obtained. In consideration of this possibility, the data have been reexamined using values of T_m three degrees higher than are listed in Table I. When the appropriate plots are then made, straight lines are obtained as illustrated by the dotted lines of Figs. 8 and 9. Thus, within the probable experimental uncertainty of locating T_m, the mechanisms proposed explain the very marked variation of the crystallization rate with temperature. A distinction between disk or spherical type nuclei is not possible, however.

As we have pointed out previously, the temperature coefficient of the nucleation incubation time τ^* must differ markedly from the observed temperature coefficient. It follows that if τ^* were appreciable compared to the observed τ_i the concurrence with the temperature-dependent term in Eqs. (16) and (17) and (18) should not have been obtained.

Although the data for the various pip-10's appear to fall on different lines within the experimental error of locating their equilibrium melting temperature, all lines can be superimposed by shifting the T_m's slightly. Thus, in this temperature interval the crystallization rate is not appreciably affected by the molecular weight; in particular there is no effect of melt viscosity over the thousandfold range covered by these samples, when comparison is made at the same value of $T_m^2/T(\Delta T)^2$. The results obtained for the rate constant k_d^d of duplicate samples of pip-10A and of pip-10B are presented in Fig. 9. The open triangles represent the results for one of the pip-10B samples, the solid triangles the results for the other sample. The open inverted and the solid inverted triangles are the results for the two samples of pip-10A. Excellent agreement is obtained for duplicate samples of both polymers.[37]

The molecular parameters which determine the temperature variation of the rate constant in the temperature interval investigated are the crystal nuclei-liquid interfacial free energies, the bulk heat of fusion per unit volume of repeating unit, and the equilibrium melting temperature. Values of the latter two parameters were obtained from melting measurements. The heats of fusion per unit volume of repeating unit, Δh_u, are easily calculated from ΔH_u, the heat of fusion per mole of repeating unit and the molar volume at T_m. Values of the interfacial free energies can be estimated from the slopes of the straight lines of Figs. 8 and 9. Only the results for the case of disk nucleation and spherical growth are presented, since it was found that the variation in σ_S for the different nucleation and growth mechanisms was a mere 10 percent. The products $\sigma_S^2\sigma_E$ which are obtained directly from the slopes treated according to Eq. (18) are given in column six of Table I. The values of σ_S given in the next to the last column were calculated by assuming that for each of the polymers σ_E/σ_S equals 0.5. The values obtained for σ_S are not unreasonable, although perhaps lower than might have been anticipated.[38] Following Turnbull's[39] analyses of crystallization kinetics in metal droplets, there is given in the last column of Table I the ratio of $\sigma_G/\Delta H_u$, where σ_G is the surface free energy of a mole of repeating units on a surface one unit thick. For metals this ratio is found to fall in the range 0.3 to 0.5; the values for the polymers studied are somewhat lower.

Over a wider range of temperature, and in particular at temperatures far below T_m, it is no longer permissible to disregard the variation in the growth rate G with temperature. By reference to Eq. (16), (17), or (18),

[35] D. E. Roberts and L. Mandelkern, Phys. Rev. **91**, 225 (1953).
[36] Mandelkern, Roberts, and Quinn, unpublished results on the melting transition of polymethylene.

[37] Plots of log k_d^d against $T_m/T\Delta T$ (not shown) are satisfactorily linear for all five polymers when the values of T_m given in Table I are used. An interpretation in accord with this temperature coefficient consists in the formation of disk nuclei of fixed thickness which possess only lateral surface energy. The formation of a nucleus of this type is highly arbitrary, and in view of the very real uncertainties in the correct value of T_m, the straight lines that are found in this case are probably fortuitous.
[38] D. G. Thomas and L. A. K. Staveley, J. Chem. Soc. 4569 (1952).
[39] D. Turnbull, J. Appl. Phys. **21**, 1022 (1950).

it will be evident that the second term in the exponential becomes quite small when ΔT is large. The first term will dominate and the temperature coefficient will be reversed, i.e., the rate of crystallization will decrease with decreasing temperature because of the decreased mobility. It follows from the foregoing theory that a maximum should appear in the rate at intermediate temperatures. The occurrence of such a maximum is well established in all examples where crystallization rates have been investigated over a wide temperature range.[7,9]

To a better approximation we observe that the "apparent" activation energy E_D for viscous flow in practice is not independent of the temperature, but may be approximated by $E_D=E_D^0/T^n$, where n ranges from one to five depending on the polymer.[32,33] Introducing this expression into Eq. (18), for example, we obtain

$$k_s^d=(\pi/3\delta)G_0^3A_0 \left[-4E_D^0/\mathbf{k}T^{n+1}-\kappa T_m^2/(\Delta T)^2\mathbf{k}T\right], \quad (19)$$

where $\kappa=8\pi\sigma_S^2\sigma_E/(\Delta h_u)^2$. The maximum rate will occur at the temperature T_{max} for which the exponent is a minimum. This temperature is defined by

$$4(n+1)E_D^0/\kappa = T_m^2(3T_{max}-T_m)T_{max}^n/(T_m-T_{max})^3. \quad (20)$$

The temperature at which the rate of crystallization is a maximum depends on the nucleation parameter κ, the activation energy of viscous flow, and on its temperature dependence. These parameters, unfortunately are not available for either natural rubber or poly-(ethylene terephthalate) but an estimate of T_{max}/T_m can be made by assuming $n=0$, $E_D=5$–7×10^{-13} erg/molecule, and taking values of κ ranging from 1–2×10^{-15} erg which are typical of the polymers in the present investigation. It is then found that T_{max}/T_m is in the range 0.8–0.9, which agrees with the experimental observation of approximately 0.85.[7,9] The theory developed to describe the crystallization kinetics at temperatures in the neighborhood of T_m is at least qualitatively valid over the whole temperature range.

It is planned shortly[14] to discuss an extension of these ideas to the description and explanation of crystallization kinetics in polymer-diluent systems.

X-Ray and Infrared Studies on the Extent of Crystallization of Polymers*

J. B. Nichols
Chemical Department, Experimental Station, E. I. du Pont de Nemours and Company, Wilmington, Delaware

(Received January 12, 1954)

The current x-ray and infrared techniques for estimation of the crystalline-amorphous ratio in polymers are surveyed. Representative data are assembled for some of the common crystallizing polymers. Advantages and limitations of the methods are pointed out.

INTRODUCTION

CRYSTALLINITY and orientation, together with molecular weight, are the most important factors determining the mechanical properties of polymers. Until a few years ago, estimation of crystallinity was purely qualitative and usually consisted of comparing the relative sharpness of the x-ray diagrams or determining the density of the polymer. In recent years, thanks to the work of Gehman and Field, Bunn, Goppel and Arlman, Hermans, Matthews and Peiser, and Bryant, Krimm, and Tobolsky,[1] quantitative x-ray data may be obtained. Still more recently, from the pioneer work of Sutherland, Thompson, Mochel and Hall, and Cobbs,[2] infrared techniques are available for reliable estimates of crystallinity for a number of common polymers.

This review attempts to summarize recent x-ray and infrared work on crystallinity with attention to orientation only where it enters as a complicating factor. X-ray and infrared techniques in use for determination of crystallinity are discussed, their advantages and limitations are pointed out, and data on representative polymers, especially polyethylene, are given. This polymer has a definite amorphous band in the x-ray pattern which is sufficiently separated from the crystalline spacings so that it may be resolved with some assurance. On the other hand, polymers like nylon, polyethylene terephthalate, and cellulose have a continuous background of unordered material underlying the x-ray pattern. They are therefore more difficult to resolve into two fractions, and the significance of the measurement is perhaps somewhat less definite.

In assigning a scale of crystallinity, the most difficult problem is to obtain a truly amorphous sample with which to compare the sample under investigation. This problem has been handled in various ways, for example by use of x-ray or infrared data on molten polyethylene or a liquid hydrocarbon as a standard of reference for polyethylene; rubber above its crystalline melting point for rubber; and "amorphous" cellulose prepared by ball-milling for cellulose. If the data are standardized with density values there is again the uncertainty of the exact value of the density of the amorphous polymer, although that of the ideal crystal may be calculated from the x-ray unit cell dimensions if they have been determined.

X-RAY METHODS

X-ray runs that have been carried out under carefully standardized conditions as to constancy of x-ray intensity, constancy of mass in the beam, and random orientation of sample, give considerable information on the percent crystallinity (usually amorphous content is determined), and size and shape of the crystallites from the broadening of the x-ray lines representing the different spacings. Even where orientation is present, pictures may be taken in more than one direction, or the sample may be rotated to average out the effect on the intensities of the lines. Figure 1 gives x-ray patterns of a nearly amorphous polyethylene terephthalate compared with a highly crystalline sample. Here the blurred ring actually is the overlapping of several rings so that there is really no distinct amorphous band but rather it underlies all of the rings. Figure 2, showing the diffraction patterns of three polyethylenes of different amorphous contents, obtained with an x-ray spectrometer by Bryant,[3] gives an example of the amorphous peak partly

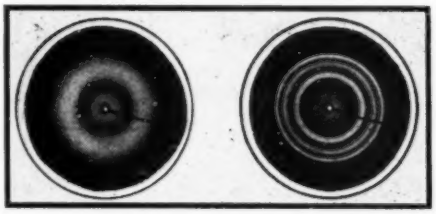


FIG. 1. X-ray patterns of polyethylene terephthalate.

* Presented at the Symposium on Crystallization of Polymers at the Division of High-Polymer Physics, American Physical Society Meeting, Durham, North Carolina, March 27, 1953.

[1] S. D. Gehman and J. E. Field, J. Appl. Phys. **10**, 564 (1939). C. W. Bunn and J. C. Alcock, Trans. Faraday Soc. **41**, 317 (1945). J. M. Goppel and J. J. Arlman, Appl. Sci. Research **A1**, 462 (1947–49). P. H. Hermans and A. Weidinger, J. Polymer Sci. **4**, 135 (1949). Matthews, Peiser, and Richards, Acta Cryst. **2**, 85 (1949). Bryant, Tordella, and Pierce, Jr., Paper read at the 118th National Meeting of the American Chemical Society, Chicago, September 3–8, 1950. S. Krimm and A. V. Tobolsky, J. Polymer Sci. **7**, 57 (1951).

[2] G. B. B. M. Sutherland and A. V. Jones, Discussions Faraday Soc. **9**, 274, 281 (1950). H. W. Thompson, J. Chem. Soc. **1947**, 287; Thompson, Nicholson, and Short, Discussions Faraday Soc. **9**, 222 (1950). W. E. Mochel and M. B. Hall, J. Am. Chem. Soc. **71**, 4082 (1949). W. H. Cobbs, Jr., and R. L. Burton, J. Polymer Sci. **10**, 275 (1953).

[3] Bryant, Tordella, and Pierce, reference 1.

resolved from the main bands. The two peaks with indices 110 and 200 are crystalline reflections and represent interferences due to the principal side spacings in the polyethylene structure. The shoulder on the low-angle side of the 110 peak is part of the broad diffraction band characteristic of the amorphous regions in the polymers.

The method of measurement of amorphous content by x-ray diffraction is in principle a comparison of the relative amounts of energy diffracted by the crystalline and by the essentially amorphous regions in the polymer and rests on the assumption of equivalence in scattering power of the two regions for equal masses. Figure 3 indicates the method of resolving the peaks for a polyethylene pattern.[3] This is based on the symmetrical character of the peaks. The background is drawn in to follow the general background on each side of the pattern as judged by the background for a liquid octa-

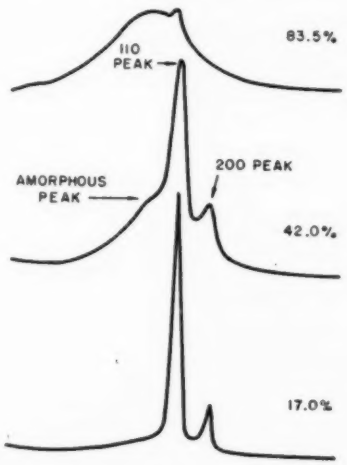


FIG. 2. X-ray spectrometer records showing the influence of amorphous content on the diffraction pattern for polyethylene (Bryant, see reference 1).

decane. The background correction is the most uncertain feature of the method and has resulted in rather different results by different investigators. The center of the amorphous band is taken at a spacing of 4.50 A at room temperature. The area is then resolved into three symmetrical peaks.

The three areas are planimetered and are corrected for atomic scattering factors, absorption, temperature, and diffraction angle. The combined correction factors calculated for these positions are 0.69, 1.00, and 1.43, respectively. The measured areas after correction represent relative intensities I_A, I_{110}, and I_{200}. Since the intensity of reflection, whether from the crystalline or the amorphous regions in polyethylene, has been found to be closely proportional to mass, the amorphous content A of the polymer on a percent basis is expressed by the relation,

$$A(\%) = 100 \cdot \frac{I_A}{I_A + I_{100} + I_{200}} = 100 \cdot \frac{0.69a}{0.69a + b + 1.43c}.$$

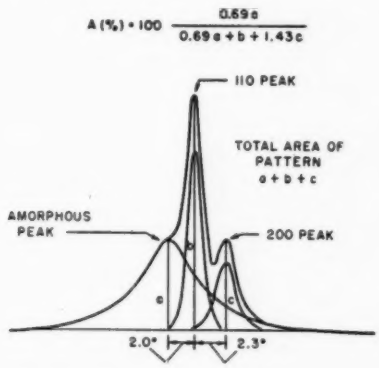


FIG. 3. Method of x-ray analysis of amorphous content in a polyethylene sample (Bryant, see reference 1).

Krimm and Tobolsky[4] compare the area under the amorphous peak with that for molten polyethylene corrected to the same mass as the whole sample. Hermans and Weidinger[5] in their work on cellulose (Fig. 4) compare the amorphous content of the sample with that of milled amorphous cellulose over the angular range from 4° to 42° after correcting for incoherent and air scattering.

ORIENTATION CONTRIBUTION

When a preferred orientation occurs in the x-ray pattern, the method outlined above for obtaining amorphous content will be considerably in error. Figure 5 gives an example of a polyethylene film in which considerable planar orientation is present. The normal ratio of intensities of the two principal peaks should be about 2.4. In the top diagram taken through the sheet, the intensity ratio I_{110}/I_{200} is about 6, whereas in the lower pattern through the edge of the sheet the intensity ratio is only 1.3. If these two patterns are weighted according to the multiplicity of the 110 and

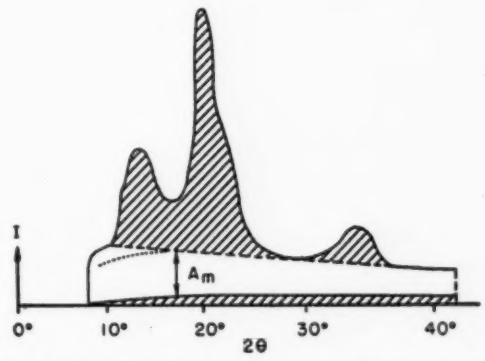


FIG. 4. Radial photometer curve of the x-ray pattern of a ramie pellet. Upper hatched area relates to the crystalline fraction. Bottom hatched area, incoherent scatter. Dotted line, correction for air scatter. Height A_m, relative measure of disordered portion (Hermans, see reference 5).

[4] S. Krimm and A. V. Tobolsky, reference 1.

[5] P. H. Hermans and A. Weidinger, reference 1: Bryant, Tordella, and Pierce, reference 1.

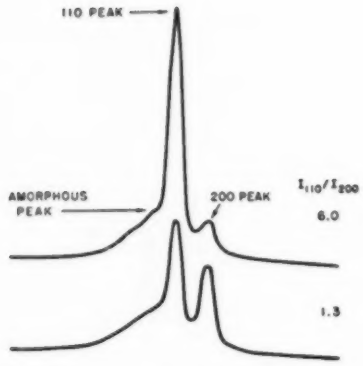


FIG. 5. Effect of planar orientation on the x-ray pattern of polyethylene. Upper curve X perpendicular plane of sheet. Lower curve X parallel edge of sheet (Bryant, see reference 1).

200 planes, namely 2:1, and averaged, the amorphous content may be obtained as outlined before with relatively small error.

With a yarn, if the sample is rotated, Ingersoll[6] found the orientation to be fairly well averaged out, as is evident from Fig. 6 showing the pattern from an unstretched rayon yarn and that of a stretched, oriented sample which has been rotated.

INFRARED MEASUREMENTS

Infrared spectra of polymers, subjected to physical treatments which change the crystalline content, undergo easily measurable changes in intensity of certain absorption bands. If bands can be found which are sensitive to amorphous content and to crystalline content, an absolute measure of crystallinity may be made.

If, on the other hand, a given polymer shows only sensitivity to changes in crystallinity then it is necessary to compare the absorbance, that is, the logarithm of the reciprocal of the transmission of the given band, with

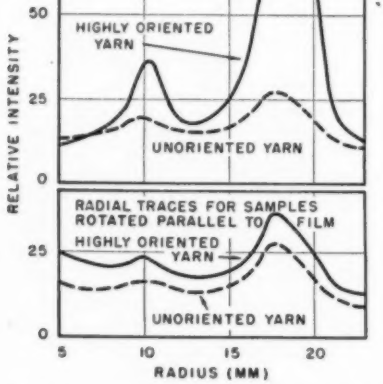


FIG. 6. Effect of rotation on the x-ray patterns of highly oriented and unoriented rayon yarns (Ingersoll, see reference 6).

[6] H. W. Ingersoll, J. Appl. Phys. **17**, 924 (1946).

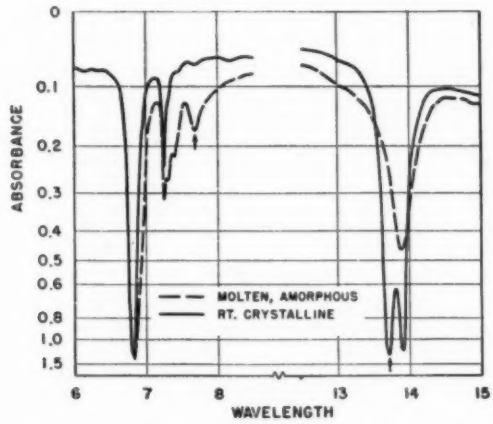


FIG. 7. Infrared spectra of amorphous and crystalline polyethylene wax (Hall, see reference 7).

the density or other suitable property of the polymer. Even though this is a relative method, it has its merits because the infrared data are less subject to fluctuations than density which is affected by a variable content of voids and other extraneous features in the sample. This is especially true if an internal standard for thickness control is present such as the CH band at 3.4 microns for rubber, nylon, and polyethylene terephthalate.

Figures 7 and 8 give examples of the reciprocal relationships between amorphous and crystalline content shown by polyethylene[7] and polyhydroxyacetic ester.[8] On the other hand, the spectra for neoprene polymerized at −40°C,[9] polyethylene terephthalate, and nylon 66 show only crystallization-sensitive bands (Figs. 9, 10, and 11).

Where both the amorphous and crystalline bands are present, the crystalline-amorphous ratio C/A is evi-

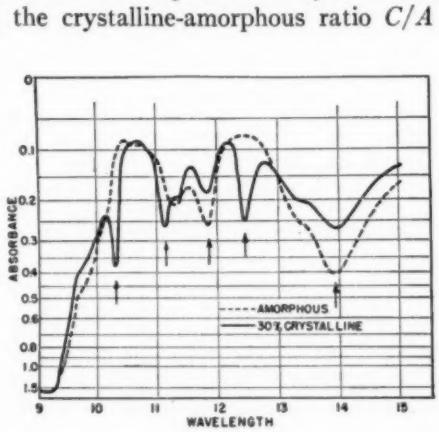


FIG. 8. Infrared spectra of amorphous and partly crystalline polyhydroxyacetic ester (Marston, see reference 8).

[7] Courtesy of M. B. Hall, Chemical Department, E. I. du Pont de Nemours and Company.

[8] Courtesy of A. L. Marston, Textile Fibers Department, E. I. du Pont de Nemours and Company.

[9] J. T. Maynard and M. B. Hall, paper presented at the 123rd National Meeting of the American Chemical Society, Los Angeles, March 15–19, 1953, J. Polymer Science **12**, 235 (1954).

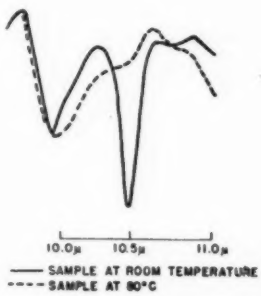


FIG. 9. Infrared spectra of melted and frozen polychloroprene polymerized at −40°C (Maynard and Hall, see reference 9).

dently equal to

$$K \times \frac{\text{observed absorbance crystalline band}}{\text{observed absorbance amorphous band}},$$

where K is the ratio of the absorbances for 100 percent amorphous to 100 percent crystalline samples.

This constant may be evaluated by the method of successive approximations from data for highly crystalline and highly amorphous samples. These absorbances refer, of course, to the same thickness of sample; however, the observed ratio is independent of thickness since the absorbances are determined for two bands in the same sample.

Where only a single sensitive band is available, the crystallinity values must be calibrated by some other method such as density, as mentioned before. If orientation affects the intensity of the band, a separation of the effects can frequently be made by locating a band which is only dichroic. For example, Sutherland[10] noted that the 8.87 and the 11.9 micron bands in stretched latex rubber increased in strength on cooling to −140°C, whereas the 6.02 μ dichroic band showed no change. By use of these bands for separation of the dichroic

TABLE I. Crystallization-sensitive infrared bands.

Polymer	Amorphous	Crystalline	Reference
Polyethylene	7.7μ	13.65μ	Richards[a]
Polyethylene terephthalate	...	7.4, 10.3, 11.8	Hall[b] Cobbs[c]
Rubber	Near 11.9?	8.87, 11.9	Sutherland[d]
Neoprene	...	10.5	Maynard[e] Hall
Nylon 66	1.48	1.53, 10.7	Hall[b] Ellis[f]
Polyvinyl alcohol	1.42	1.52, 1.6	Zhurkov[g]
Hydroxyacetic polyester	11.8, 13.9	10.3, 12.4	Marston[h]

[a] See reference 20. See also G. D. Buckley and N. H. Way, J. Chem. Soc. 3701 (1952).
[b] See reference 7.
[c] See reference 2.
[d] G. B. B. M. Sutherland, Discussions Faraday Soc. 9, 274 (1950).
[e] See reference 9.
[f] L. Glatt and J. W. Ellis, J. Chem. Phys. 15, 880 (1947); 16, 551 (1948).
[g] S. N. Zhurkov and B. Ya. Lovin, Akad. Nauk S. S. S. R., Doklady 67, 89 (1949).
[h] See reference 8.

[10] G. B. B. M. Sutherland and A. V. Jones, reference 2.

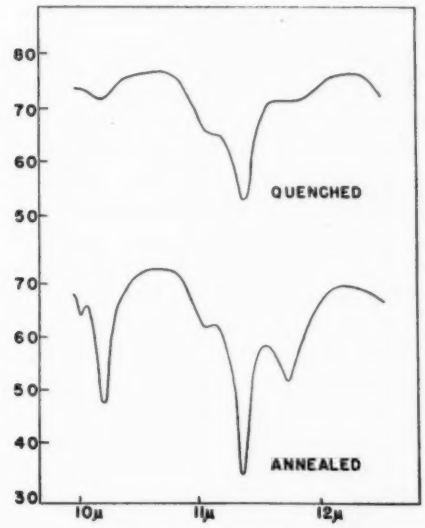


FIG. 10. Infrared spectra of quenched and annealed polyethylene terephthalate.

ratio from crystallization he was able to obtain a crystallinity figure for highly stretched rubber, approximately 40 percent, similar to that of Goppel and Arlman.[11]

Table I gives a summary of crystallization-sensitive bands for different polymers in the infrared with a few which are also sensitive to amorphous content. Some of the bands are sensitive to orientation as well but at least can be used for random samples without correction.

ADVANTAGES AND LIMITATIONS

Table II lists some of the principal advantages and limitations of the methods in use.

INDIVIDUAL POLYMERS

The data obtained on individual polymers by the above methods, especially on polyethylene, cellulose, and rubber, are reviewed below.

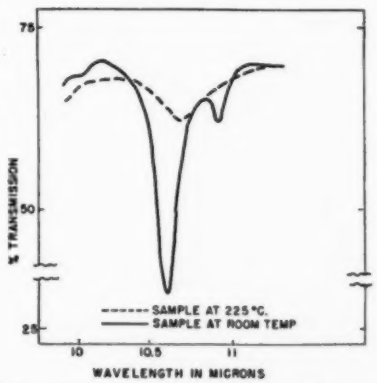


FIG. 11. Infrared spectra of crystalline and amorphous nylon 66.

[11] J. J. Arlman, Appl. Sci. Research A1, 347 (1947–49); J. M. Goppel and J. J. Arlman, reference 1.

1. Polyethylene

Polyethylene has probably received the most attention because of the definite amorphous band which can be resolved on the x-ray diagram. The studies of Krimm and Tobolsky[4] and of Bryant, Tordella, and Pierce[3] are probably the freest from objection because they were carried out with x-ray spectrometers but they show some discrepancies.

Figure 12 gives data obtained by Bryant and his coworkers on a number of samples of shock cooled and annealed polyethylene covering an amorphous range from about 15 to 70 percent. There is evidently a nearly linear relationship between density and x-ray amorphous content. The dashed line extrapolates to a value of 0.805 for the amorphous density. However, the full line may be a better representation, based on additive volumes. Data of other investigators are indicated on the plot. Those of Matthews, Peiser, and Richards[12] are given in the dotted curve. The lower amorphous content found by them was due solely to their choice of a narrower angular scattering region, which resulted in a higher background correction. The amorphous contents obtained by Raine, Richards, and Ryder[13] by heat content were also too low because their heat of fusion was too low, 56 calories per gram as compared with about 60 calories, determined later by Parks and Mosely.[14] The values obtained by Hunter and Oakes[15] by density measurements, and by Hermans and Weidinger[16] by x-ray measurements of the ratio of intensities (at $2\theta = 7$ to 11°) of the amorphous band to that of molten polyethylene, are in fair agreement.

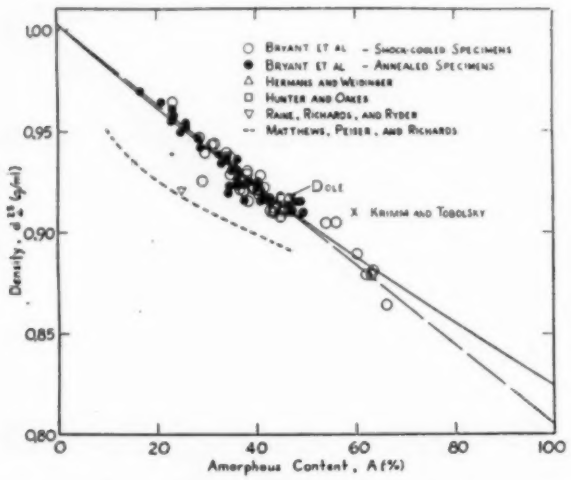


FIG. 12. Relation between density and amorphous content for various polyethylenes (Bryant, see reference 1).

[12] Matthews, Peiser, and Richards, reference 1; see also Trillat, Barbezat, and Delalande, Compt. rend. 231, 853 (1950).
[13] Raine, Richards, and Ryder, Trans. Faraday Soc. 41, 56 (1945).
[14] G. S. Parks and J. R. Mosely, J. Chem. Phys. 17, 691 (1949).
[15] E. Hunter and W. G. Oakes, Trans. Faraday Soc. 41, 48 (1945).
[16] P. H. Hermans and A. Weidinger, J. Polymer Sci. 4, 709 (1949).

TABLE II. Comparison of methods for crystallinity determination.

Method	Advantage	Limitation
X-ray	Permits determination of both orientation and crystallization	Uncertain background
	Gives estimate of crystallite size	Rigorous control of method required
Infrared	Rapid, less overlap of bands	Orientation interferes
	Internal standard for mass	
	Covers wide crystallinity range	
Density	Rapid, simple	Holes. Attached air
	No uncertain background correction	Poor wetting

The similar sample run by Krimm and Tobolsky[4] appeared to have a somewhat higher amorphous content relative to its density, about 60 percent *versus* 50 percent for Bryant's sample of the same density. On the other hand, a recent estimate of Dole[17] of the crystallinity by heat content gave a value closer to Bryant's curve. Dole's value of the heat of fusion of about 64 at room temperature may be too high; so, if $H_f = 60$, his sample would fall on the curve.

Table III points out the main difference between Bryant's and Krimm's techniques. Both used x-ray spectrometers so that photographic uncertainties were avoided, but Krimm used crystal-monochromatized Cu-Kα radiation, whereas Bryant used heavily nickel-filtered radiation. Krimm calculated and subtracted the incoherent scattering and air scattering, whereas Bryant made a single background correction "consistent with the general background" shown by a liquid hydrocarbon octadecane. This feature may be one of the important differences.

Krimm used molten polyethylene at 120°C as his amorphous standard and took the ratio of the actual amorphous area to that of the molten polyethylene to obtain his amorphous content. This seems to be a considerable extrapolation. On the other hand, Bryant took the ratio of the amorphous area to the total of the amorphous-crystalline scattering and thus had an internally consistent set of data. In addition, he made synthetic mixtures of crystalline polyethylene in liquid octadecane, and confirmed the identity of the relative scattering efficiencies of the amorphous and the crystalline material. Krimm included scattering at smaller angles down to 2° (range 2° to 30°), whereas Bryant covered the range from 10° to 30°.

In comparing crystallinity obtained by density with that by x-ray methods, Krimm took a value of 0.86 density for 100 percent amorphous material, obtained by extrapolating the density of molten polyethylene to 24° by means of the temperature coefficient of expansion. On the other hand, Bryant used a value of 0.805 to 0.825. It seems unlikely that the amorphous density at room temperature is higher than 0.82 or 0.83. In

[17] Dole, Hettinger, Larson, and Wettington, J. Chem. Phys. 20, 781 (1952).

TABLE III. Comparison of x-ray techniques—polyethylene.

	Bryant, Tordella, and Pierce[a]	Krimm and Tobolsky[b]
Apparatus	Spectrometer	Spectrometer
Character x-rays	Ni filtered	Crystal monochromatized
Angular range	10–30°	2–30°
Background corr.	Liquid octadecane reference	Air and calc'd incoherent scatter subtracted
Amorphous reference	Ratio crystl. to amorph. area	Molten polyethylene 120° (two records)
Density used	ca. 0.82 from additive volumes	0.86 extrap. from molten polyethylene

[a] See reference 1.
[b] See reference 1.

fact, Hermans' data[18] from liquid paraffin and polyethylene indicate that there may be more incipient order in liquid paraffin than in "amorphous" polyethylene. Matthews and Peiser[1] calculated a density of 0.78 from the average interchain spacing of 5.2 A on the assumption of a roughly hexagonal arrangement of the chain molecules in the amorphous regions of polyethylene. The results indicate that there is still some uncertainty but that the trend has been to estimate a higher amorphous content in polyethylene as more careful investigations have been made.

Crystallinity Versus Temperature

Krimm and Tobolsky[4] have determined the decrease in crystallinity with temperature by their x-ray methods and Dole[17] and others[13] have used heat capacity to determine similar relationships. Figure 13 gives the findings together with data obtained by density methods.

The weight fraction of CH_2 groups in the crystalline form x was obtained from the relation $V_{obs.}=xV_{cryst.}+(1-x)V_{amor.}$. The specific volumes $V_{cryst.}$ and $V_{amor.}$ were obtained from the x-ray crystal structure estimate of 1.0 at 25° and the volume coefficient of expansion by extrapolation of the density-temperature curve for liquid polyethylene. They considered the higher crystallinity indicated by the density method at higher temperatures to be due to fringe areas of higher density rather than to strictly random amorphous regions.

The crystallinity from the heat capacity measurements was obtained from the ratio of the difference between the heat content of the liquid polyethylene extrapolated below 115° and the measured heat content divided by the heat of fusion; Dole considered it to lie between 62 and 66 calories per gram depending on the temperature.

Crystallization Versus Branching

As mentioned earlier, the intensities of many of the bands in the infrared spectrum of polyethylene change when the sample is melted or crystallized. The band at 13.65 microns increases with crystallinity, whereas that at 7.67 μ increases with disorder. A measurement of the intensity of these bands in a sample of solid polyethylene gives a direct measure of the proportions of crystalline and amorphous regions. Rugg[19] states that estimation of crystallinity from the 13.65 μ band is affected by stretching but this orientation effect could probably be eliminated by taking spectra both through the plane and the edge of the plane, as Bryant did in his x-ray studies.[3]

Properties depending on slight movements of portions of the sample relative to one another are particularly dependent on crystallinity. Low molecular weight polyethylene is more dense and more crystalline down to about 2000 molecular weight. The more short chain branching, the less dense the sample and the less crystalline. Richards and co-workers[20] in England have prepared polymethylene from diazomethane which is essentially linear with only 5 percent amorphous content. With four methyl groups per hundred carbon atoms, this value has risen to 45 percent as indicated in Fig. 14. Here the amorphous content has been determined by a combination of x-ray, infrared, and density measurements by Richards.

Crystallinity Versus Elongation

Krimm and Tobolsky found for low elongation of the order of 20 percent that the amorphous regions were still random and that crystallinity could be determined. On the other hand, for elongations of the order of 100 percent to 500 percent, there was a definite orientation of the amorphous regions as well as the crystalline regions so that they were not able to estimate the effect of high elongations on crystallinity. However, it seems likely that x-ray patterns taken in two directions or

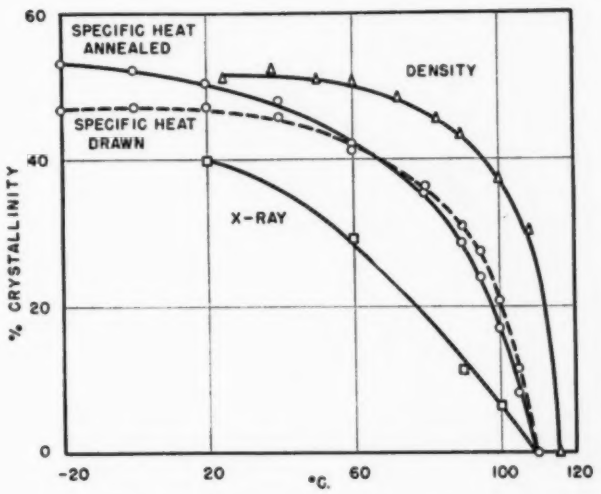


FIG. 13. Effect of temperature on the crystallinity of polyethylene (Dole, see reference 17).

[18] P. H. Hermans and A. Weidinger, J. Polymer Sci. 4, 714 (1949).

[19] Rugg, Smith, and Atkinson, J. Polymer Sci. 9, 579 (1952).
[20] R. B. Richards, J. Appl. Chem. 1, 370 (1951).

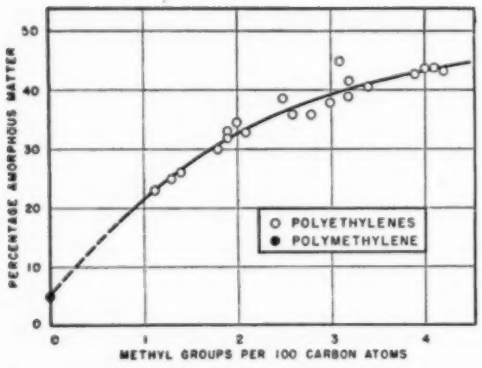


FIG. 14. Effect of branching on amorphous content of polyethylene (Richards, see reference 20).

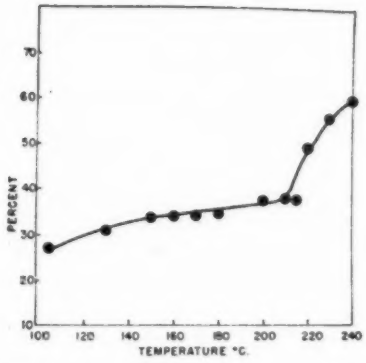


FIG. 16. Effect of heat treatment on the crystallinity (infrared) of polyethylene terephthalate film (Cobbs, see reference 2).

rotation of the sample would have helped to average out this orientation.

They did, however, estimate the change in average crystallite size as a function of drawing and found it to drop in the direction perpendicular to the 110 planes from 140 A to 104 A for 100 percent stretch; for 500 percent stretch average crystallite size dropped to 59 A.

2. Cellulosic Materials

Over a period of years the crystallinity of cellulosic materials has received much attention. The work of Hermans,[21] Mark,[22] Ingersoll,[6] and Sisson[23] has done much to clarify the picture. The most striking conclusion that Hermans has reached is that various native celluloses have nearly the same range of crystallinity, approximately 60–70 percent, whereas regenerated celluloses are of the order of 40 percent crystalline. It was difficult to establish an x-ray scale until Hess and his co-workers[24] found that an essentially amorphous material could be made simply by ball-milling cellulose. In this manner a product of 8–10 percent crystallinity could be obtained.

Ingersoll[6] showed that the so-called degree of lateral order could be varied independently of the orientation in agreement with Kratky's findings[25] that the degree of crystallization is not affected by orientation alone. Mark's data[26] determined on various types of cellulosic materials by x-ray, density, and isotope exchange methods are given in Table IV. Data on water vapor absorption also give essentially the same results.

Size of Crystallites

Hermans[27] attempted to estimate the size of crystallites from the integrated intensities of the x-ray diffraction peaks and their half widths. He found that in native cellulose the intensity ratios of the line breadths show little variation and the crystallites seem to possess about the same average size, shape, and degree of perfection. In regenerated fibers, however, the shape and

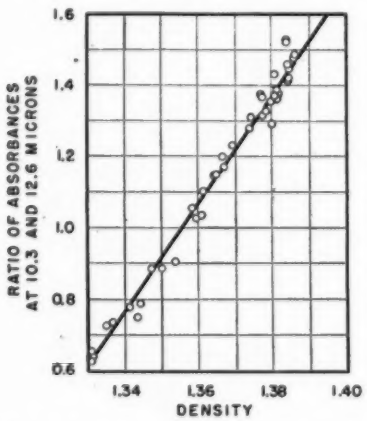


FIG. 15. Relation between density and infrared absorbance for unoriented polyethylene terephthalate film (Cobbs, see reference 2).

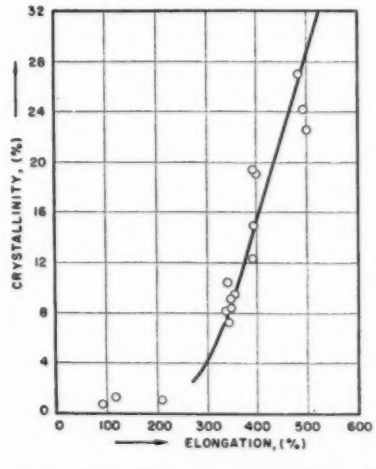


FIG. 17. Effect of stretching on the crystallinity of compounded natural rubber (Goppel and Arlman, see reference 33).

[21] P. H. Hermans, *Physics and Chemistry of Cellulose Fibers* (Elsevier, Amsterdam, 1949).
[22] H. Mark, J. Phys. Chem. **44**, 764 (1940).
[23] W. A. Sisson, Chem. Rev. **26**, 196 (1940).
[24] K. Hess, H. Kiessig, and J. Gundermann, Z. physik. Chem. **B49**, 64 (1941).
[25] O. Kratky and A. Sekora, Kolloid-Z. **108**, 169 (1944).
[26] R. Brill and H. Mark, Z. Elektrochem. **55**, 202 (1951).
[27] P. H. Hermans, J. Polymer Sci. **4**, 145 (1949).

perfection of the crystallites vary considerably. The crystallites are ribbonlike with the 101 plane in the ribbon. The differences among various rayons seem to be in thickness of the ribbon, whereas the other directions show little change. Therefore, the ratio of the intensity of the $10\bar{1}$–002 combination to that of the 101 peak gives a means of differentiation. Apparently increased thickness of the lamellae decreases the total number as judged by the weakening of the 101 interference.

3. Polyethylene Terephthalate

Data by Cobbs[28] for polyethylene terephthalate show a linear relation between density and infrared absorbance of the 10.3 μ band (Fig. 15). Here the absorbance of the 12.6 μ band was used as an internal standard. As with polyethylene, there is some uncertainty as to the exact value of the amorphous density, 1.30–1.33.

For polyethylene, examples were given of the effect of temperature on the extent of crystallization. A similar curve is not available for polyethylene terephthalate, but Cobbs has determined the effect of heat-treating a polyethylene terephthalate film. In this case the film was subjected to a certain temperature and then allowed to cool before determining the density and the infrared spectrum at 30°C. Figure 16 gives the relation obtained between crystallinity and temperature of heat treating. A large increase in crystallinity has occurred in samples heated above 210°, apparently a point where the structure loosens sufficiently for reorganization to occur.

4. Other Polymers

Natural rubber has also received considerable attention including x-ray examination by Goppel and Arlman[29] and Hermans,[30] infrared examination by Sutherland,[31] and dilatometric examination by Wood and co-workers.[32] The improved technique of Arlman[33] over that used by Goppel[34] gave values for aged, smoked, and crepe rubbers of the order of 30–40 percent crystallinity which verified Wood's estimates based on a density of 1.00 for crystalline rubber and 0.91 for amorphous rubber. Vulcanized rubber shows an extremely rapid increase in crystallization above an elongation of 300 percent. Arlman and Goppel[35] find essentially the same extent of crystallization as a function of elongation from x-ray measurements and from the use of Flory's theoretical relations. Figure 17 shows the increase in crystallization of the vulcanized rubber as a function of elongation.

To summarize the data available on common polymers, Table V is given with estimates of the crystallinity of various polymers by x-ray, infrared, and density methods where possible. The density of the ideal crystal calculated from x-ray data and that of the amorphous material are also given if available. The crystallinity figures given do not necessarily represent a characteristic range for the polymers except possibly for polyethylene. The value obtained depends usually on the physical treatment given the polymer during processing.

TABLE IV. Crystallinity of cellulosic materials by various methods.

Material	% Crystallinity X-ray	Density	Deuterium
Ramie	70	70	75
Mercerized ramie	47	40–50	...
Cotton	65–75	60–70	75–82
Mercerized cotton	40–50	40	...
Wood pulp	60–65	50–55	63
Viscose rayons	40–45	40	40–50

TABLE V. Summary of crystallinity of polymers.

Polymer	Method[a]	Density Amorphous	Density Crystalline	Percent crystallinity
Polyethylene	X, IR, D, H	0.78 0.86	1.00	30–90
Polytetrafluoroethylene	X	<2.23	2.42	<87
Polychlorotrifluoroethylene	D	...	2.25±	<90[b]
Neoprene	X, IR, D	1.24	1.35	10–40
Rubber	X, IR, D	0.91	1.00	<50
Polyethylene terephthalate	I, IR, D	1.30 1.33	1.46	<75
Polyvinylidene chloride	X, D	1.66	1.95	<75[c]
Polyvinylidene chloride/polyvinyl chloride		...	...	19–67[c]
Cellulose	X	...	1.58	8–70
Nylon 66	X	1.09	1.24±0.04	30–50[d]

[a] X = x-ray, IR = infrared, D = density, H = heat content.
[b] F. P. Price, J. Chem. Phys. **19**, 973 (1951).
[c] R. F. Boyer and R. S. Spencer, J. Appl. Phys. **15**, 398 (1944).
[d] P. H. Hermans and A. Weidinger, J. Polymer Sci. **4**, 718 (1949).

[28] W. H. Cobbs, Jr., and R. L. Burton, reference 2.
[29] J. M. Goppel and J. J. Arlman, reference 1.
[30] P. H. Hermans and A. Weidinger, J. Polymer Sci. **4**, 713 (1949).
[31] G. B. B. M. Sutherland, Discussions Faraday Soc. **9**, 274 (1950).
[32] L. A. Wood in *Advances in Colloid Science* (Interscience Publishers, Inc., New York, 1946), Vol. II, p. 57.
[33] J. J. Arlman, Appl. Sci. Research **A1**, 347 (1947–1949).
[34] J. M. Goppel, Appl. Sci. Research **A1**, 3, 18 (1947–1949).
[35] J. J. Arlman and J. M. Goppel, Appl. Sci. Research **A2**, 1 (1949–1951).

An Infrared Study of the Crystallization of Polyethylene*

F. P. Reding and Alexander Brown
Research and Development Department, Carbide and Carbon Chemicals Company, South Charleston, West Virginia

(Received January 12, 1954)

The melting point, temperature for the onset of crystallization, and the isothermal rate of crystallization of polyethylene all decrease on increasing the extent of branching of the polymer chains. In the molecular weight range studied the melting point is independent of the molecular weight. The temperature for the onset of crystallization at any given rate of cooling appears to increase as the molecular weight is decreased. The isothermal rate of crystallization becomes much greater as the crystallization temperature is lowered, however, at a given temperature the rate is affected only slightly by a change in molecular weight and is primarily a function of the degree of branching.

POLYETHYLENE is a crystalline polymer which, in its most common form, was first shown by Hunter and Oakes[1] to be approximately 50 to 60 percent crystalline. Usually the molecules are not linear, but rather contain many side chains or branches.[2]

Since the branch junctures cannot fit into the crystal lattice and the terminal methyl groups of the branches would cause a discontinuity in the lattice, the crystalline content and consequently many of the physical properties of polyethylene are determined by the degree of branching.[3]

The present study was undertaken to determine the effect of varying the branching and the molecular weight upon the melting point, temperature for the onset of crystallization and the isothermal rate of crystallization.

EXPERIMENTAL

The 13.7-micron band of the infrared spectrum, which arises from the crystalline phase,[4] was used for all measurements. To determine the melting and crystallization temperatures, a 0.002-inch thick film of the polymer was sealed between rocksalt plates and a thermocouple was inserted between the plates into the body of the polymer. The sample was then placed in an oven mounted on a spectrometer and the change in intensity of the 13.7-micron band was recorded on heating or cooling.

To determine the isothermal rates of crystallization a procedure was used similar to that used by Cobbs and Burton[5] to measure the rate of crystallization of polyethylene terephthalate. Two ovens, one directly over the other, were mounted on the spectrometer. The ovens were separately controlled; the upper oven was heated to 150°C and the lower one heated to any desired temperature below the melting point. A 0.004-inch thick polyethylene film was clamped in a thin metal frame and placed in the upper oven. After three minutes at 150°C the sample frame was dropped into the lower oven so that it was in the path of the infrared radiation. Since the temperature of the lower oven was controlled by a rapidly flowing stream of air, the melted sample could be cooled quickly and the desired temperature below the melting point could be attained before crystallization began. At temperatures considerably below the melting point, however, a sample would remain melted for only a very short period of time. For this reason the isothermal rate of crystallization experiments could not be carried out below 85°C.

In all the above experiments a calibrated copper-constantan thermocouple was used. The determinations of melting and crystallization temperatures were reproductible within ±0.2°C. The isothermal rate of crystallization experiments were carried out with the thermocouple very close to but not quite touching the film. The error in these temperature measurements was approximately ±0.5°C.

RESULTS AND DISCUSSION

Melting Point

The melting points, viscosity molecular weights, and methyl contents (degree of branching) for the series of polyethylenes investigated are shown in Table I. The melting point was taken as the temperature at which the last evidence of 13.7-micron absorption disappeared. Sample D-1 is the only exception to the rule that increasing branching decreases the melting point. In this case, however, the double bond and oxygen content was fairly high. If it is assumed that each double bond and oxygen containing group has the same effect as a methyl group, then the effective methyl content is 1.2 percent and the melting point of D-1 is not out of line with the other samples. The melting points were not affected by changing the rate of heating from 1/5 to 1°C per minute. Thus it appears that over a wide range of molecular weights the melting point of polyethylene is dependent only on the extent of branching. This is in

* Presented at the Symposium on Crystallization of Polymers at the Division of High-Polymer Physics, American Physical Society Meeting, Durham, North Carolina, March 27, 1953.

[1] E. Hunter and W. G. Oakes, Trans. Faraday Soc. **41**, 49 (1945).

[2] J. J. Fox and E. A. Martin, Proc. Roy. Soc. (London) **A175**, 208 (1940).

[3] R. B. Richards, J. Appl. Chem. **1**, 370 (1951).

[4] King, Hainer, and McMahon, J. Appl. Phys. **20**, 559 (1949).

[5] W. H. Cobbs and R. L. Burton, J. Polymer Sci. **10**, 275 (1953).

TABLE I.

Sample	Chain branching % CH_3	Molecular weight	Melting point T_m°C	Crystallization temperature T_c Cooling rate 0.5°C/min T_c	T_m-T_c	Cooling rate 3.0°C/min T_c	T_m-T_c	Cooling rate 11°C/min T_c	T_m-T_c
Polymethylene	Nil	High	128.4	122.1	6.3	114.9	13.5	112.5	15.9
D-1	0.3(1.2)	2600	106.8	102.3	4.5	...	...	96.3	10.5
D-2	0.7	29 000	108.4	102.0	6.4	98.2	10.2	97.3	11.1
D-3	1.8	22 000	106.5	98.5	8.0	95.1	11.4	94.5	12.0
D-4	2.5	22 000	105.5	97.0	8.5	93.5	12.0	92.9	12.6
D-5	3.0	22 000	105.0	96.3	8.7	90.9	14.1	90.1	14.9
D-6	3.5	13 000	104.2	96.0	8.2	92.5	11.7	90.0	14.2

substantial agreement with the results of other investigators.[3,6]

This result is just what one would expect in light of Flory's[7] theory of crystallization in high polymers. In all probability the longest methylene chains between branches determine the melting point. Since this is a function of the methyl-to-methylene ratio and not the molecular weight, the melting point should be a function of the extent of branching alone.

It is interesting to note that if methods other than the disappearance of the 13.7-micron band are used to determine the melting points, the values obtained are several degrees higher. This indicates that the crystalline dipole coupling which gives rise to the 13.7-micron band is destroyed several degrees before the last form of crystalline order disappears.

Crystallization Temperature

The temperatures for the onset of crystallization (hereafter referred to as T_c) for various rates of cooling the melt are listed in Table I. It is seen that T_c is decreased markedly below the melting point by the slow rates of cooling and then is decreased only slightly more by the more rapid rate of cooling. In general, T_c decreases with increasing branching at a given rate of cooling, but a large part of this decrease in T_c results from the fact that the melting points of the more highly branched polymers are lower. The use of T_m-T_c takes out the effect of the lower melting points on T_c. A close inspection of the T_m-T_c columns in Table I shows that increasing branching causes a greater decrease in T_c for a given rate of cooling and a given molecular weight than can be accounted for by the decrease in T_m alone.

T_m-T_c divided by the rate of cooling gives a measure of the nucleation time, the time required for stable crystal nuclei to form after the polymer is cooled below the melting point. The nucleation times for the various polymers and rates of cooling are listed in Table II. The nucleation time is high for the slowest rate of cooling then decreases sharply with increasing rate of cooling.

A comparison of D-1 and D-6 with the other samples shows that molecular weight has a small effect. The data indicate that decreasing the molecular weight increases T_c and therefore decreases T_m-T_c and the nucleation time.

Isothermal Rate of Crystallization

Isothermal rates of crystallization of the four samples listed in Table III were investigated. The rates were measured for each sample at several different temperatures, but because of the difficulty in setting the temperature accurately to any predetermined point, the rates for the different samples were not run at exactly the same temperatures. The fraction of the sample which crystallized per minute was calculated from the experimental curves.† Although the method of calculating the rate is arbitrary, the calculated values correspond very nearly to the maximum values of the crystallization rates which are the slopes of the S-shaped experimental curves at the inflection points. The calcu-

TABLE II.

Sample	Chain branching % CH_3	Molecular weight	Nucleation time in minutes for different rates of cooling 0.5°C/min	3.0°C/min	11°C/min
Polymethylene	Nil	High	12.6	4.5	1.45
D-1	0.3(1.2)	2600	9.0	...	0.96
D-2	0.7	29 000	12.8	3.4	1.01
D-3	1.8	22 000	16.0	3.8	1.09
D-4	2.5	22 000	17.0	4.0	1.15
D-5	3.0	22 000	17.4	4.7	1.35
D-6	3.5	13 000	16.4	3.9	1.29

TABLE III.

Sample	Chain branching % CH_3	Molecular weight	Rate of crystallization in % crystallinity/min 100°C	90°C
D-2	0.7	29 000	15	180[a]
D-7	0.8	22 000	12	500
D-3	1.8	22 000	2.9	125
D-5	3.0	22 000	0.8	35

[a] The reversal of the relative rates of crystallization of D-2 and D-7 in going from 100° to 90°C results from their different molecular weights.

† This calculation was made by dividing one-fourth of the equilibrium percent crystallinity of the sample at the temperature in question by the time $(t_{1/2}-t_{1/4})$, where $t_{1/2}$ is the time required for the polymer to crystallize to one-half its equilibrium value and $t_{1/4}$ is the time required for the polymer to crystallize to one-fourth its equilibrium value. The equilibrium percent crystallinity at any temperature was determined by correlating the intensity of the 13.7-micron band and the density.

[6] V. K. Ueberreiter and H. Orthmann, Kolloid-Z. 128, 1 (1952).
[7] P. J. Flory, J. Chem. Phys. 17, 223 (1949).

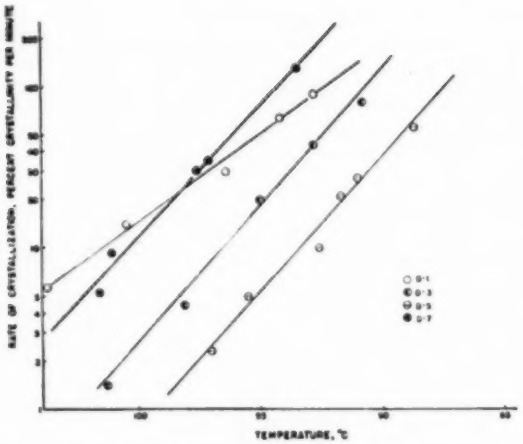


FIG. 1. Isothermal rate of crystallization *vs* temperature.

lated rates on log scale are plotted against temperature in Fig. 1.

The plots are linear in the temperature range shown within the accuracy of the data and curves for the polymers of the same molecular weight, D-3, D-5, and D-7, are parallel. Little basic significance can be attributed to the linearity of the curves since at the melting point the rate must be zero and the logarithm of the rate minus infinity and at temperatures lower than those measured the rate must decrease.

It is seen in the figure that the rate of crystallization increases very rapidly with decreasing temperature. For D-3, D-5, and D-7 the rate doubles for every 1.8°C temperature drop, and for D-2, the higher molecular weight polymer, the rate doubles for every 2.8°C temperature drop. This increase in rate with decreasing temperature is what one would expect in light of the theories of nucleation[8] and the kinetics of crystallization of high polymers.[9] Just below the melting point crystal nuclei are formed slowly, but as the temperature is lowered nuclei are formed much more rapidly. Consequently, even though the rate of crystallization around the individual nuclei is lower, there are so many more growing nuclei at the lower temperatures that the over-all rate of crystallization is greater.

At some temperature below the melting point the lower chain mobility and consequent decrease in the rate of crystallization around a nucleus must cause the over-all rate of crystallization to level off and finally decrease with decreasing temperature. The nucleation time for polyethylene, however, is so short that this temperature range could not be reached before crystallization began with the present experimental technique.

A second factor which has a marked effect on the rate of crystallization is the degree of branching. Decreasing the degree of branching from 3 percent methyl in D-5 to 1.8 percent in D-3 to 0.8 percent in D-7 causes an increase in their relative rates of crystallization from 1 to 3.5 to 16, respectively, at any given temperature. D-2, the higher molecular weight polymer, also crystallizes much more rapidly than the more highly branched samples. It is seen from the different slopes that molecular weight has an effect on the rate of crystallization but the data are insufficient for any analysis. For comparison the rates of crystallization of the four samples at 100°C and 90°C are listed in Table III.

Avrami[10] has treated the kinetics of phase transformation and has derived the equation

$$A = \exp - Bt^k,$$

where A is the uncrystallized fraction at time t, and B and k are independent of time. Thus a plot of log$t$ *vs* loglog $1/A$ should be a straight line with slope k. Such plots were made from several of the isothermal rate curves and proved to be linear for about 90 percent of the course of a crystallization. Avrami further predicted that the slope k should be between 1 and 4. The plots for the different samples and temperatures all had slopes of approximately two. The linearity of the plots over most of the course of crystallization indicates that the nucleation is homogeneous and the differences between the rates of crystallization of the different samples results from the differences in branching and not from the presence of greater or smaller numbers of heterogeneously nucleating bodies in the polymers.

[8] Fisher, Hollomon, and Turnbull, J. Appl. Phys. **19**, 775 (1948).

[9] T. Alfrey, Polymer Bull. **1**, 40 (1945).

[10] M. Avrami, J. Chem. Phys. **7**, 1103 (1939); **8**, 212 (1940).

Crystallization in Butadiene-Styrene Copolymers*†

LAWRENCE A. WOOD
National Bureau of Standards, Washington, D. C.
(Received January 12, 1954)

Butadiene-styrene copolymers contain the following units: *trans*-1,4-polybutadiene, *cis*-1,4-polybutadiene, 1,2-polybutadiene, and bound styrene. Of these the first is the only type which has been shown to form crystals. By the use of published data on volume change arising from crystallization (largely due to Meyer) and on structure as a function of polymerization temperature (largely due to Hampton) it is found that crystallization in emulsion polymers has been observed only when the *trans*-1,4-polybutadiene content has been greater than about 58 percent, becoming more complete the more this value is exceeded. From this observation one can then draw the following conclusions, all of which are in accordance with direct experimentation:

(1) Crystallization is not observed if the polymerization temperature is above 60°C.

(2) For polymerizations at 50°C a small amount (2 to 6 percent) of bound styrene inhibits crystallization completely.

(3) For polymerizations at 5°C the limit is at about 15 to 18 percent bound styrene content.

(4) At −40°C (the lowest polymerization temperature normally utilized) this limit is at about 30 percent bound styrene.

I. INTRODUCTION

UNTIL about 1948 it was quite doubtful whether or not any crystallization effects could be observed in butadiene polymers and copolymers.[1] Since then conclusive evidence has been obtained by different methods in several laboratories[2–7] that crystallinity can be obtained under favorable conditions in such polymers if prepared at low temperatures. The present paper aims to utilize published experimental data from several sources to outline more clearly the structural considerations and the conditions of observation under which crystallization can be expected to be observed.

Beu, Reynolds, Fryling, and McMurry[2] produced crystals in polybutadiene prepared at low temperatures and by x-ray studies showed them to have an identity period of 5.1 A, corresponding to the *trans*-1,4 configuration. This work confirmed earlier surmises[1,8] that the crystals observed are those of the material in the *trans*-1,4 configuration contained in a mixture with the *cis*-1,4 configuration and the 1,2 configuration. It will be evident from observations to be discussed later that none of the butadiene polymers which have thus far been prepared and investigated contain the *cis*-1,4 configuration as the major constituent. Consequently, it is not surprising that crystals corresponding to the *cis*-1,4 configuration have never been observed.

II. CHARACTERISTICS OF CRYSTALLIZATION

Crystallization in polymers is conveniently studied by following volume changes in liquid-filled dilatometers[3,4,6,9]. In unstretched polybutadiene crystallization is found to be fairly rapid in comparison with that of natural rubber, corresponding to the simpler unit of polybutadiene. Lucas and co-workers[3] showed that the change in volume arising from crystallization is greater the lower the temperature of polymerization (see for example Fig. 3 of their paper). They also showed that in copolymers of butadiene and styrene the volume change associated with crystallization is less the higher the styrene content of the copolymer (see for example Fig. 2 of their paper).

The primary effect of a change in crystallization temperature on the other hand is a change in the rate of crystallization, according to detailed studies on natural rubber.[9] Small amounts of cross-linking apparently produce a similar change in rate.[10] In both cases there is relatively little effect on the volume change after extended times.

The rate of crystallization of butadiene polymers and copolymers, like that of natural rubber, passes through a maximum as the temperature of crystallization is decreased. Calorimetric data of Furukawa and others[7] extending down to −258°C show crystallization oc-

* This work was supported in part by the U. S. Office of Naval Research and the Office of the Quartermaster General.

† Presented at the Symposium on Crystallization of Polymers at the Division of High-Polymer Physics, American Physical Society Meeting, Durham, North Carolina, March 27, 1953.

[1] L. A. Wood in book H. Mark and G. S. Whitby, *Advances in Colloid Science, Volume II, Rubber* (Interscience Publishers, Inc., New York, 1946).

[2] Beu, Reynolds, Fryling, and McMurry, J. Polymer Sci. **3**, 465–479 (1948); reprinted in Rubber Chem. Tech. **22**, 356–369 (1949).

[3] Lucas, Johnson, Wakefield, and Johnson, Ind. Eng. Chem. **41**, 1629 (1949).

[4] A. W. Meyer, Ind. Eng. Chem. **41**, 1570 (1949).

[5] H. N. Campbell and M. D. Allen, Ind. Eng. Chem. **43**, 413 (1951).

[6] J. R. Beatty and B. M. G. Zwicker, Ind. Eng. Chem. **44**, 742 (1952).

[7] Furukawa, McCoskey, and King, J. Research Natl. Bur. Standards **50**, 357 (1953) Research Paper RP2425.

[8] H. Mark in book R. E. Burk and O. Grummitt, *The Chemistry of Large Molecules* (Interscience Publishers, Inc., New York, 1943).

[9] L. A. Wood and N. Bekkedahl, J. Appl. Phys. **17**, 362 (1946); J. Research Natl. Bur. Standards **36**, 489 (1946), RP1718; Rubber Chem. Tech. **19**, 1145 (1946).

[10] N. Bekkedahl and L. A. Wood, Ind. Eng. Chem. **33**, 381 (1941); Rubber Chem. Tech. **14**, 347 (1941).

curring only in the range −73° to +22°C. The optimum temperature for maximum rate of crystallization, by analogy with that for natural rubber, may be expected to occur near or somewhat below the middle of this range.

Modern theories of crystallization in polymers have led to the expectation that the optimum would depend on the melting temperature and that the latter would be depressed systematically by increasing amounts of non-crystallizing material. Consequently under ideal conditions one would make separate determinations of the optimum temperature for each polymer, and then make measurements of volume change as a function of structure or composition after infinite time at the optimum temperature for the particular polymer. Optimum temperatures have not been determined as a function of structure or composition, and data on volume changes at these temperatures are not available. However, the crystallization temperature −45°C used in the studies of Meyer[4] is probably fairly near the optimum for any of the polymers considered here. It is difficult to know whether the conclusions of Meyer would have been altered appreciably by extending the observations to longer times of crystallization than the 200 minutes (3⅓ hours) usually used in his work.

III. STRUCTURAL CONSIDERATIONS

Butadiene-styrene copolymers are made up of four types of units in varying amounts: *trans*-1,4-polybutadiene, *cis*-1,4-polybutadiene, 1,2-polybutadiene, and polystyrene. In polybutadiene itself, obviously, only the first three types of units appear. In several recent studies[6,11–17] infrared techniques have been applied to determining the relative proportions of the different types of unit. While the numerical values obtained differ slightly all the results agree in showing that emulsion polymers of butadiene have a 1,2-polybutadiene content which varies only slightly with temperature of polymerization, increasing from about 16 percent at $T_p=-40$°C to about 20 percent for T_p from 100° to 270°C. The *trans*-1,4-polybutadiene content is found to be of the order of 80 percent at the lowest polymerization temperatures, decreasing to about 50 percent at $T_p=100$°C and to about 45 percent at $T_p=270$°C. Correspondingly, the *cis*-1,4-polybutadiene content increases from about 4 percent at $T_p=40$°C to about 30 percent at $T_p=100$°C. The rate of increase above this temperature is much less, with a value of about 35 percent reported for $T_p=270$°C. The values just given are, to a first approximation, independent of the polymerization recipe for emulsion polymers. Quite different values are of course obtained for alkali-catalyzed polymers.

Simple equations relate the quantities considered. If B_T, B_C, B_V, and B are respectively the *trans*-1,4, the *cis*-1,4, the 1,2 (vinyl), and total polybutadiene content while S is the bound styrene content, all expressed in weight percent, then

$$B_T+B_C+B_V+S=100 \tag{1}$$

and

$$F_T=\frac{100B_T}{B}=\frac{100B_T}{100-S}, \tag{2}$$

where F_T is the percent of the polybutadiene units in the *trans*-1,4 configuration.

The data of Hampton[12] have been selected for use in the quantitative calculations reported in the present paper. Only slight numerical changes would have resulted from the selection of data from the other observers. All of Hampton's data for the fraction of polybutadiene units in the *trans*-1,4 configuration for both polybutadienes and copolymers were assembled and plotted as a function of the temperature of polymerization. To a reasonable approximation both sets of data could be represented by a straight line with the equation:

$$F_T=70-0.21T_p, \tag{3}$$

where T_p is the temperature of polymerization (in degrees Celsius). The equation is regarded as satisfactory for temperatures of polymerization up to about 100°C.

Hampton's results, including copolymers containing not more than 27 percent bound styrene, showed no significant differences in the relative proportions of the different polybutadiene units with changes in bound styrene content. Recent studies by Foster and Binder[16] of copolymers containing up to about 86 percent bound styrene show an increase of F_T with increasing amounts of bound styrene. The increase is marked above about 30 percent bound styrene. Below this value the experimental uncertainty is such that the increase is not clearly evident. Since the present paper is concerned only with the lower values of bound styrene F_T is taken as independent of the styrene content. A correction term could be readily introduced if necessary.

Combining Eqs. (2) and (3) one obtains

$$100B_T=(100-S)(70-0.21T_p) \tag{4}$$

as the equation relating *trans*-1,4-polybutadiene content to styrene content and polymerization temperature.

[11] E. J. Hart and A. W. Meyer, J. Am. Chem. Soc. **71**, 1980–1985 (1949); Rubber Chem. Tech. **23**, 98–106 (1950).
[12] R. R. Hampton, Anal. Chem. **21**, 923 (1949).
[13] J. L. Binder, unpublished reports to Office of Synthetic Rubber, Reconstruction Finance Corporation, dated September 17, 1951, March 4, 1953, and March 30, 1953.
[14] W. S. Richardson, "The microstructure of diene polymers. I. Polyisoprenes and polybutadienes prepared at high temperatures," unpublished report to Office of Synthetic Rubber, Reconstruction Finance Corporation dated September 1952.
[15] F. E. Condon, J. Polymer Sci. **11**, 139 (1953).
[16] F. C. Foster and J. L. Binder, J. Am. Chem. Soc. **75**, 2910 (1953).
[17] A. I. Medalia and H. H. Freedman, J. Am. Chem. Soc. **75**, 4790 (1953).

IV. CRYSTALLIZATION AND STRUCTURE

The volume change ΔV on holding samples at $-45°C$ for 200 minutes ($3\frac{1}{3}$ hours) has been observed by Meyer[4] using liquid-filled dilatometers. This quantity, which is proportional to the fraction of the polymer in the crystalline state, is plotted as ordinate in Fig. 1. The abscissa is the *trans*-1,4-polybutadiene content as calculated from Eq. (4) utilizing Meyer's reported temperature of polymerization and a value of styrene content calculated from the reported charging ratios. The points denoted by a cross were obtained with samples of polybutadiene polymerized at different temperatures while the other points were obtained in other sets of experiments in which the styrene content of a copolymer was systematically varied at different fixed polymerization temperatures.

The fact that all of the values fall, within experimental error, on a single curve indicates that B_T, the *trans*-1,4-polybutadiene content, is the controlling factor in determining the extent of the crystallization. The relative proportions of the noncrystallizing units have no effect on the extent of the crystallization—at least to a first approximation. It is only their total amount that counts. If the amount of crystallization depended directly on the styrene content or on the temperature of polymerization rather than upon the resulting *trans*-1,4-polybutadiene content, one would have several different curves, whereas it can be seen that the data are represented adequately by a single curve.

The straight dashed line in Fig. 1 is drawn between the origin and a point corresponding to $\Delta V = 3.5$ percent for $B_T = 100$ percent. It may be that for values of B_T above 75 percent the noncrystallizing types of units act as mere diluents in reducing the volume change in a manner proportional to their concentration. Below 75 percent however it is clear that this is not true. In this region the reduction is more nearly proportional to the fourth power of the concentration.

According to the curve in Fig. 1, if the trans-1,4-polybutadiene content falls below about 58 percent the volume change in the given time at $-45°C$ is zero and

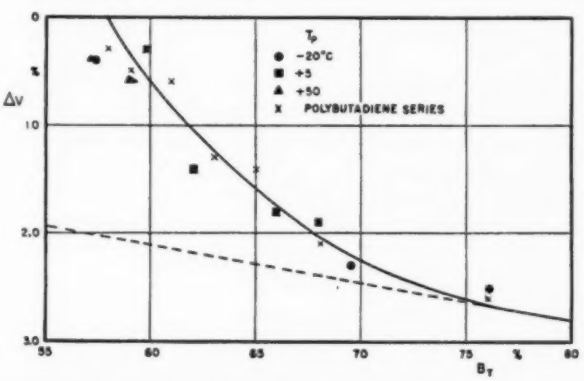


FIG. 1. Volume change as a function of *trans*-1,4-polybutadiene content.

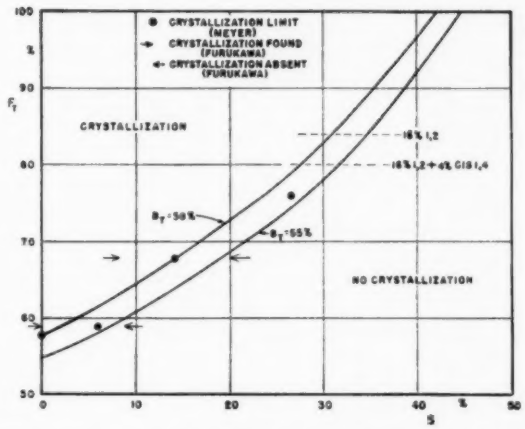


FIG. 2. Fraction of polybutadiene units in *trans*-1,4 configuration as a function of bound styrene content for two fixed values of *trans*-1,4-polybutadiene content. An arrow pointing to the right indicates that specific heat measurements showed the presence of crystallization for the composition indicated by the tip of the arrow; an arrow pointing to the left indicates the absence of crystallization for the composition indicated by the tip of the arrow.

no crystallization is observed. This is the most significant conclusion of the present paper. If the observations had been conducted at the optimum temperature for each polymer and if they had been continued to longer times it is possible that the curve would be lowered somewhat and the intercept would be reduced below 58 percent. Beatty and Zwicker,[6] studying a copolymer containing about 46 percent *trans*-1,4-polybutadiene, have reported a volume change of about 0.1 percent after 32 hours at $-64°C$. This volume change, estimated at about 5 times the minimum detectible with the dilatometers used, is much smaller than any of the changes plotted in Fig. 1 and indicates a very slight amount of crystallization at a lower level of *trans*-1,4-polybutadiene content than other observers have reported. The significance of this observation is not clear. The volume change may represent crystallization in portions of the polymer where statistical fluctuations have brought the *trans*-1,4-polybutadiene content higher than the average for the whole polymer, or the temperature of crystallization employed may be nearer the optimum for this particular polymer than those used by the other observers.

Some of the consequences of the crystallization limit just mentioned are illustrated in Fig. 2. The abscissa is the bound styrene content of the copolymer and the ordinate is the fraction of the polybutadiene units which are in the *trans*-1,4 configuration. The upper curve represents a constant content of *trans*-1,4-polybutadiene equal to 58 percent. This line may be regarded as dividing the region of the graph into two areas, in one of which crystallization is possible and in the other of which crystallization has not been observed. Another curve representing a constant *trans*-1,4-polybutadiene content of 55 percent is also shown to correspond to the possibility of a somewhat lower limit if observations

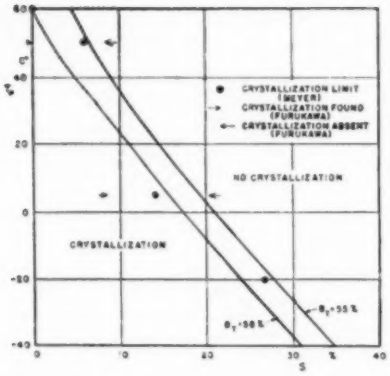


FIG. 3. Polymerization temperature as a function of bound styrene content for two fixed values of *trans*-1,4-polybutadiene content. An arrow pointing to the right indicates that specific heat measurements showed the presence of crystallization for the composition indicated by the tip of the arrow: an arrow pointing to the left indicates the absence of crystallization for the composition indicated by the tip of the arrow.

had been extended over a longer period of time. There is in this figure no opportunity to represent the extent of the crystallization, in terms of ΔV, for example. The curves are graphical representations of Eq. (2) and are of course not dependent on any experimental observations.

The fraction of the polybutadiene units in the 1,2 configuration for the emulsion polymers which have been studied has never been reported below about 16 percent. This sets a limit shown by the upper dashed line, below which all points would be expected to lie. The lowest reported value of the fraction of polybutadiene units in the *cis*-1,4 configuration is about 4 percent. With the polymerization systems and temperatures now in use, therefore, experimental points will lie below the lower dashed horizontal line.

The four points in Fig. 2 which are indicated by circles represent values for the crystallization limit computed from Meyer's observations.[4] These have been already utilized in Fig. 1. The four arrows in Fig. 2 represent experimental observations of Furukawa and co-workers.[7] An arrow pointing to the right indicates that specific heat measurements showed the presence of crystallization; an arrow pointing to the left indicates the absence of any crystallization. The values of S defined by the tips of the arrows were obtained by direct measurement of carbon-hydrogen ratio. The values of F_T defined by the tips of the arrows were obtained from the reported polymerization temperatures by the use of Eq. (3). As expected, the lines representing the crystallization limit are found to lie between the tips of the arrows.

Figure 3 is similar to Fig. 2 in many respects except that the ordinate is now the temperature of polymerization. Because of this fact the curves are necessarily based on experimental observations of F_T as a function of polymerization temperature. Hampton's observations as represented by Eq. (3) have been used. The experimental points, now independent of Hampton's observations, are again taken from the data of Meyer and Furukawa. This form of representation is perhaps the most convenient for practical use.

From Fig. 3 one draws the following significant conclusions:

(1) Crystallization is not observed if the polymerization temperature is above 60°C.

(2) For polymerization at 50°C a small amount (2 to 6 percent) of bound styrene inhibits crystallization completely.

(3) For polymerizations at 5°C the limit is at about 15 to 18 percent bound styrene content.

(4) At the lowest polymerization temperatures normally utilized this limit is at about 30 percent bound styrene.

Direct experimental observation[3–7] is in general accord with these conclusions.

Regular Contributed Original Research

Modified Luneberg Lens

A. S. GUTMAN*
Air Force Cambridge Research Center, Cambridge, Massachusetts
(Received August 24, 1953)

Spherically symmetric optical instruments, and their desirability for microwaves, are discussed. Application of Hamiltonian optics in analyzing the rays in the Luneberg lens has led to formula of a new lens that has spherical symmetry and a focus near its center. A cylindrical version of this new lens has been designed for operation in the TE_{10} mode.

MODIFIED LUNEBERG LENS

IN this paper it will be shown that it is theoretically possible to build a large-diameter big-aperture stationary antenna, free of spherical aberration or coma, with which a pencil beam can be scanned over the entire volume of space by moving a point source over a small-diameter sphere. Such a device has long been desired for radar application but its achievement has hitherto been held impossible. It is, as a matter of fact, impossible if the focus is in a medium of dielectric constant 1, but with a high dielectric constant at the focus,it is possible to produce a small feed circle, as this paper will show. A practical version of a cylindrical antenna structure, based on this theory, is included.

The wavelengths used in radar are much longer than those in optics, and instruments for directing the radiation, therefore, have to be correspondingly larger. Rotating such large instruments for scanning purposes is cumbersome, and the speed of rotation has its mechanical limitations. Since optical instruments of spherical symmetry have the same properties in all directions, the possibilities of their scope in microwave applications is of much theoretic interest.

Luneburg† has shown that a spherical dielectric having a variable index of refraction, with the index of refraction a function of the radius alone, can be designed so that it focuses a parallel beam of light into a point. A Luneburg lens of radius a has an index of refraction

$$n=(2-(r/a^2)^{\frac{1}{2}},$$

and the property that any point source of light at its surface $r=a$ produces a parallel beam with an aperture $2a$ on the opposite side of the lens. It is pertinent that the index of refraction in the Luneburg lens is 1 at $r=a$, so the lens is, therefore, matched to free space.

The Luneberg lens, being symmetric, has the practical advantage of being much more adaptable to rotation than other focusing structures but it still requires either rotation of the point source around the outside diameter of the structure or rotation of the entire structure. This requirement could be avoided if a spherically symmetric focusing medium had its focus near the center of symmetry. It will be shown in this paper that a spherically symmetric lens, whose focus is on the inside and whose full aperture can be used, is mathematically feasible and its cylindrical version practically possible.

Luneberg and other investigators have used Fermat's principle and the calculus of variations. I have attacked this problem by means of Hamiltonian optics (corpuscular theory of light).

Consider a set of wavefronts as surfaces

$$\zeta(x,y,z,t)=0,$$

representing a hypersurface in four-dimensional space; and consider a three-dimensional surface, which the radiation has reached at time t, as

$$\psi(x,y,z)=\text{const},$$

where the constant is a function of time t and the dimension is length^2/time; and

$$\psi_x^2+\psi_y^2+\psi_z^2=c^2n^2,$$

where

$$\psi_x=\frac{\partial\psi}{\partial x},$$

$$\psi_y=\frac{\partial\psi}{\partial y},$$

$$\psi_z=\frac{\partial\psi}{\partial z},$$

n is the dielectric constant,
c is the velocity of light in vacuum.

The trajectories of the light rays can be described by the equations

$$x=f(\tau),$$
$$y=f(\tau),$$
$$z=f(\tau),$$

where τ is an arbitrary parameter.

* Now with Sylvania Electric Products Inc., Boston Engineering Laboratory.

† *Mathematical Theory of Optics* (mimeographed notes), R. K. Luneberg, Brown University, Providence, Rhode Island.

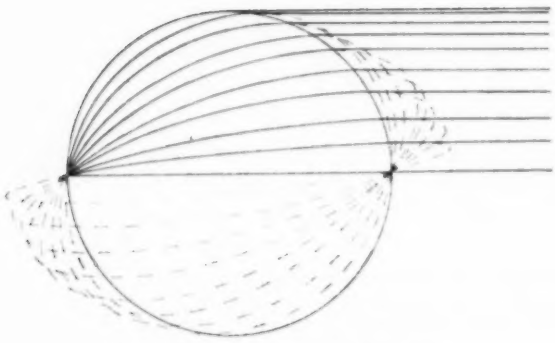


FIG. 1. Rays in Luneberg lens ($a=f$).

Light rays are orthogonal trajectories to the wavefronts and, therefore, the following differential equations have to be met:

$$\frac{dx}{d\tau}=\psi_x,$$

$$\frac{dy}{d\tau}=\psi_y,$$

$$\frac{dz}{d\tau}=\psi_z.$$

Differentiated with respect to τ,

$$\frac{d^2x}{d\tau^2}=\psi_{xx}\frac{dx}{d\tau}+\psi_{xy}\frac{dy}{d\tau}+\psi_{xz}\frac{dz}{d\tau}=\psi_{xx}\psi_x+\psi_{xy}\psi_y+\psi_{xz}\psi_z$$

$$=\frac{1}{2}\frac{\partial}{\partial x}(\psi_x^2+\psi_y^2+\psi_z^2)=c^2\frac{\partial}{\partial x}(\tfrac{1}{2}n^2).$$

Indicating the differentiation with respect to τ with a dot, and repeating this differentiation for $dx/d\tau$ and $dz/d\tau$, we get

$$\ddot{x}=c^2\frac{\partial}{\partial x}(\tfrac{1}{2}n^2),$$

$$\ddot{y}=c^2\frac{\partial}{\partial y}(\tfrac{1}{2}n^2), \qquad (1)$$

$$\ddot{z}=c^2\frac{\partial}{\partial z}(\tfrac{1}{2}n^2).$$

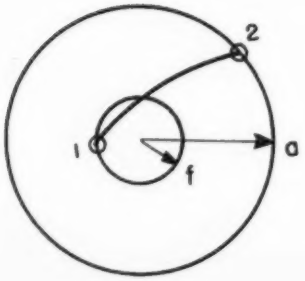


FIG. 2.

If we consider the light rays as paths of corpuscles moving in a potential field $\phi=-\frac{1}{2}n^2$ with an energy

$$\tfrac{1}{2}(\dot{x}^2+\dot{y}^2+\dot{z}^2)-\tfrac{1}{2}c^2n^2=0,$$

interpreting τ as time t, we obtain the same equations. The velocity of these corpuscles would be

$$\dot{x}^2+\dot{y}^2+\dot{z}^2=c^2n^2, \qquad (2)$$

which is greater than the velocity of light in a medium whose dielectric constant is higher than 1.

Using this corpuscular theory of light, which is valid for geometric optics, we obtain the general equation for the rays inside the Luneberg lens (see Fig. 1). With the focus at $x=a$, $y=0$, $z=0$,

$$x^2\sin^2\alpha+2xy\sin\alpha\cos\alpha+y^2(\cos^2\alpha+1)-a^2\sin^2\alpha=0.$$

These are ellipses with the parameter α. (Derivation of this equation is analogous to that of the more general case that follows and is therefore omitted here.)

It is now possible to adjust these ellipses so that the new focus falls inside the sphere. The corresponding new dielectric constant is

$$n=1/f(a^2+f^2-r^2)^{\frac{1}{2}}, \qquad (3)$$

where f is the distance from the center to the focus and a is the radius of the new lens. (See Fig. 2.) At

$$r_1=f,$$
$$n_1=a/f;$$

and at

$$r_2=a,$$
$$n_2=1.$$

A proof of the optical property of this lens follows:

Using Eqs. (1) and (3), and $r^2=x^2+y^2+z^2$, we get

$$\ddot{x}=-\frac{c^2}{f^2}x,$$

$$\ddot{y}=-\frac{c^2}{f^2}y, \qquad (4)$$

$$\ddot{z}=-\frac{c^2}{f^2}z.$$

Because the lens is symmetric, the rays are plane curves. Without loss of generality we can restrict ourselves to investigating the rays in the xy plane. The general solution of the differential Eq. (4) for the rays in the xy plane is

$$x=K_1\cos\left(\frac{c}{f}t\right)+K_2\sin\left(\frac{c}{f}t\right),$$

$$y=K_3\cos\left(\frac{c}{f}t\right)+K_4\sin\left(\frac{c}{f}t\right). \qquad (5)$$

To evaluate the four constants we use the following boundary conditions: At $t=0$,

$$x=-f,$$
$$y=0,$$
$$\dot{x}=cn_1\cos\alpha,$$
$$\dot{y}=cn_1\sin\alpha,$$
$$n_1=a/f.$$

This yields

$$K_1=f,$$
$$K_2=0,$$
$$K_3=a\cos\alpha,$$
$$K_4=a\sin\alpha,$$

and the equations of the rays become

$$x=f\cos\left(\frac{c}{f}t\right)+a\cos\alpha\sin\left(\frac{c}{f}t\right),$$

$$y=a\sin\alpha\sin\left(\frac{c}{f}t\right).$$

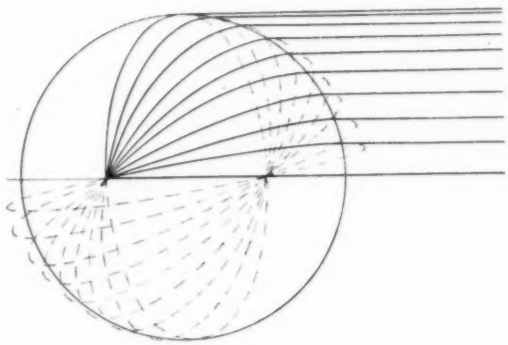


FIG. 3. Rays in modified lens ($f=\frac{1}{2}a$).

Eliminating the parameter t, we obtain

$$x^2-2xy\frac{\cos\alpha}{\sin\alpha}+y^2\left(\frac{\cos^2\alpha}{\sin^2\alpha}+\frac{f^2}{a^2\sin^2\alpha}\right)-f^2=0. \quad (6)$$

This is the equation of a family of ellipses with the parameter α. The ellipses for nine values of α, from 0° to 90°, in steps of 10°, have been plotted for

$$a=f,\quad a=\tfrac{1}{2}f,\quad a=\tfrac{1}{4}f,\quad a=\tfrac{1}{10}f,$$

in Figs. 1, 3, 4, 5, respectively. Writing Eq. (6) in the following form:

$$x=y\frac{\cos\alpha}{\sin\alpha}\pm f\left[1-\left(\frac{y}{a\sin\alpha}\right)^2\right]^{\frac{1}{2}},$$

we obtain the tangent to the ellipse parallel to the x axis by setting the quantity under the square root equal to zero, which yields

$$y=a\sin\alpha,$$
$$x=a\cos\alpha,$$
$$\overline{x^2+y^2=a^2.}$$

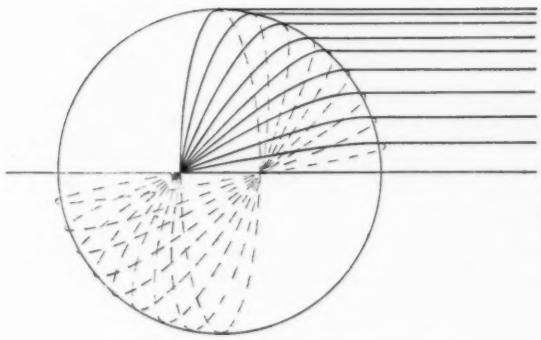


FIG. 4. Rays in modified lens ($f=\frac{1}{4}a$).

This proves that with a point source at $(-f, 0, 0)$ the medium will produce a beam of light parallel to the x axis if it is terminated at radius $r=a$.

Limiting rays are obtained for $\alpha=90°$ from Eq. (6), from which it follows that

$$\frac{x^2}{f^2}+\frac{y^2}{a^2}-1=0.$$

This is the equation of an ellipse with semi-axis a and f showing that the full aperture is utilized.

The smallest feed circle possible is determined by the dielectric constant. The highest dielectric constant is at the center ($r=0$) and this limits the minimum size of the feed circle as follows:

$$f_{\min}=\frac{a}{(n_{\max}{}^2-1)^{\frac{1}{2}}}.$$

By using dielectric materials with metal inserts it is possible to construct such a lens. The cylindrical version should be much easier to make than the spherical version. It is to be noted that the feeding wave guide has to be matched to the dielectric constant $n=a/f$ at the feed circle.

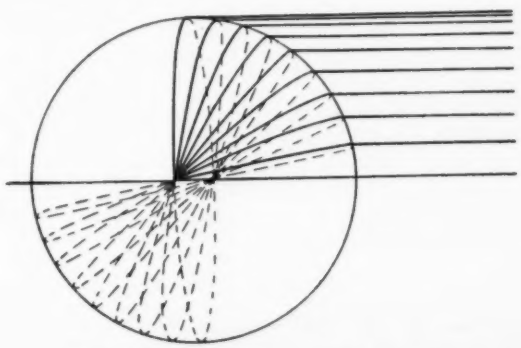


FIG. 5. Rays in modified lens ($f=1/10a$).

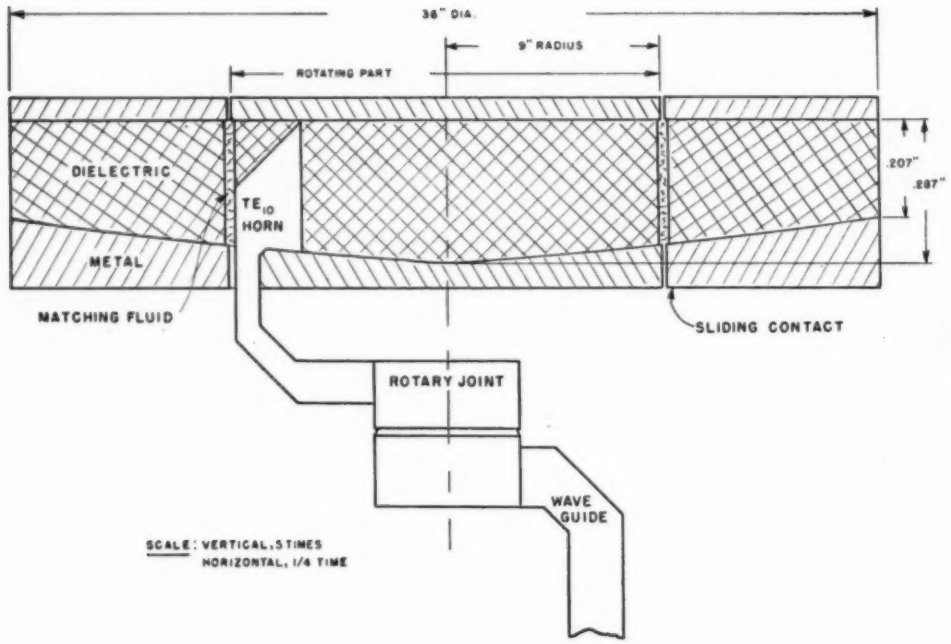


FIG. 6. Cross section schematic of cylindrical lens for TE_{10} mode.

TWO-DIMENSIONAL LENS INCORPORATING A UNIFORM DIELECTRIC, HAVING AXIAL SYMMETRY AND ITS FOCUS NEAR AXIS

A two-dimensional lens featuring a uniform dielectric, having axial symmetry and its focus near its axis, can be constructed by the technique first described by DeVore and Iams,[1] investigated by Jones,[2] and built by Peeler and Archer.[3] It is based on the fact known from wave guide theory that the effective dielectric constant between parallel plates for the TE_{10} mode is

$$n=\frac{\lambda}{\lambda_G}=\left[\epsilon'-\left(\frac{\lambda}{2u}\right)^2\right]^{\frac{1}{2}}, \tag{7}$$

where λ is the wavelength in free space, λ_G is the wavelength between plates, ϵ' is the dielectric constant of the material between the plates, and u is the spacing between the plates. This, combined with Eq. (3), yields

$$u=\frac{\lambda f}{2(\epsilon' f^2+r^2-a^2-f^2)^{\frac{1}{2}}}. \tag{8}$$

CONDITIONS FOR ϵ'

With $r=0$, in order to obtain a real value for u, it follows from Eq. (8) that we must have

$$\epsilon'>(a/f)^2+1,$$

which limits the ratio a/f to

$$a/f<(\epsilon'-1)^{\frac{1}{2}}.$$

We choose the dielectric poly 2,5-dichlorostyrene, filled with 60 percent titanium oxide, which has a dielectric constant‡ of 9.36 and a loss tangent of 0.0013. By using Eq. (3), for a wavelength of 3 cm (1.2 in.) we obtain the following formula for the spacing u:

$$u=\frac{1.2f}{2[(9.36-1)f^2-a^2-r^2]^{\frac{1}{2}}}.$$

This has been plotted in Fig. 6 for

$$a=18 \text{ in.},$$

$$f=9 \text{ in.}$$

It can be shown that this guide operates above the cutoff in the TE_{10} mode, and below the cutoff in the TE_{20} mode.

In a practical version of this lens, the inner part rotates together with the point source while the outer ring remains stationary. This necessitates a small gap between the inner and outer parts. In order to avoid reflections at the interface, this gap must be filled with a liquid of approximately the same dielectric constant as the solid. Methyl alcohol, which has a dielectric constant of 8.9 at 10^{10} cycles, may be used; its loss tangent is 0.81, but this factor should not seriously impair performance of the lens since the volume of alcohol required to fill the gap is so small.

[1] RCA Rev. IX, 729 (1948).
[2] Proc. Inst. Radio Engrs. Part III, 97, 255 (1950).
[3] NRL Report 4115 (March, 1953).

‡ *Tables of Dielectric Materials*, Vol. IV, Laboratory of Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts.

DISPERSION OF THE NEW LENS

Combining Eqs. (7) and (8), we get

$$n=\left[\epsilon'-(\lambda_0/\lambda)^2\frac{(\epsilon'-1)f^2+r^2-a^2}{f^2}\right]^{\frac{1}{2}}, \qquad (9)$$

for a lens designed for wavelength λ_0 and operated at wavelength λ. With $r=[x^2+y^2+z^2]$, this yields

$$\frac{\partial(\frac{1}{2}n^2)}{\partial x}=\left(\frac{\lambda_0}{\lambda}\right)^2\frac{x}{f^2},$$

$$\frac{\partial(\frac{1}{2}n^2)}{\partial y}=\left(\frac{\lambda_0}{\lambda}\right)^2\frac{y}{f^2},$$

which gives the following equation for the ellipses:

$$x^2=2xy\frac{\cos\alpha}{\sin\alpha}+y^2\left[\frac{\cos^2\alpha}{\sin^2\alpha}+\left[\frac{\frac{\lambda_0}{f}}{a\sin\alpha}\right]^2\right]=\left(\frac{\lambda_0}{f}\right)^2.$$

This shows that the location of the focus will move by the factor λ_0/λ as a consequence of dispersion, the size of the focal region thus corresponding to the band width of the radiation.

ACKNOWLEDGMENT

In this work, the author has had the benefit of consultation with, and cooperation of, the scientists and engineers of the Antenna Laboratory, Air Force Cambridge Research Center.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 7 JULY, 1954

Certain Transmission and Reflection Theorems*

V. TWERSKY†
Mathematics Research Group, New York University, New York, New York
(Received October 5, 1953)

The well-known relation between the total energy cross section of a scatterer and its forward scattered amplitude is extended to obtain an approximate transmission coefficient for a uniform planar distribution of parallel cylinders. An analogous reflection theorem for an arbitrary cylindrical boss on a perfectly reflecting plane is then derived; here the total cross ssction of the boss is related to the scattering amplitude in the specular direction of reflection of the plane. This is extended to obtain an approximate reflection coefficient for a uniform distribution of cylindrical bosses on a plane.

INTRODUCTION

IN this paper we treat certain theorems which are of interest in connection with the scattering and reflection of electromagnetic and acoustic waves.

In the following section, which serves primarily to introduce notation, we consider the scattering of a plane wave by a cylinder of arbitrary cross section and arbitrary physical constants. We present the well-known relation[1] that the real part of the forward scattered amplitude is proportional to the total energy cross section per unit length of cylinder, i.e., proportional to the power dissipated as heat and as scattered radiation per unit length of cylinder. We then derive a corollary of this theorem and apply it to obtain an approximate transmission coefficient for a uniform planar distribution of identical cylinders.

Next, we consider the analogous reflection problem for an arbitrary cylindrical protuberance on a smooth perfectly reflecting plane. We derive a theorem stating that the total cross section is proportional to the real part of the scattering amplitude in the direction of specular reflection of the plane. As in the transmission case, we then derive a corollary and apply it to obtain an approximate reflection coefficient for a uniform distribution of identical bosses.

THE TRANSMISSION CASE

The two-dimensional problem of the scattering of a plane wave by a cylinder of arbitrary cross section and arbitrary dielectric properties (as in Fig. 1) may be formulated as follows:

We seek a solution of the reduced wave equation $(\nabla^2+k^2)\psi=0$ for $k=2\pi/\lambda$ which satisfies prescribed boundary conditions on the cylinder, and is of the form

$$\psi=\psi_i+\psi_s \qquad (1)$$

where ψ_i is a plane wave incident at an angle α:

$$\psi_i(\theta,\alpha)=\exp[ikr\cos(\theta-\alpha)], \qquad (2)$$

* This work was performed at Washington Square College of Arts and Science, New York University and was supported in part by contract No. AF-19(122)-42 with the U. S. Air Force, Air Force Cambridge Research Center, Air Research and Development Command.

† Now with the Electronic Defense Laboratory, Mountain View, California.

[1] See for example C. H. Papas, J. Appl. Phys. **21**, 318 (1950); M. Lax, Phys. Rev. **78**, 306 (1950).

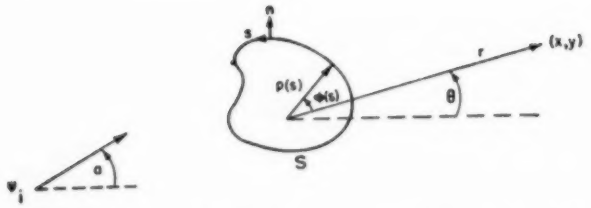


FIG. 1. Scattering of a plane wave by a cylinder. r, θ and $\rho(s)$, $\varphi(s)$ are the polar coordinates of an observation point and a point on the surface S respectively. $\mathbf{n}$ is the normal out of S and s is the coordinate along S.

and where ψ_c, the scattered wave, is an outgoing wave as $r\to\infty$; the suppressed time factor is assumed to be $\exp(-i\omega t)$. The coordinates we use are illustrated in Fig. 1.

We have in general[2]

$$\psi_c=-\int_S\left[G(r,\rho)\frac{\partial\psi(\rho)}{\partial n}-\psi(\rho)\frac{\partial G(r,\rho)}{\partial n}\right]ds \qquad (3)$$

where G is the two-dimensional free space Green's function

$$G(r,\rho)=(i/4)H_0^{(1)}(k|r-\rho|). \qquad (4)$$

For $kr\gg1$, $r\gg\rho$ we have

$$G(r,\rho)\sim(2/i\pi kr)^{\frac{1}{2}}(i/4)e^{ik[r-\rho\cos(\varphi-\theta)]}$$
$$\equiv H(r)(i/4)e^{-ik\rho\cos(\varphi-\theta)}, \qquad (4')$$

and consequently, Eq. (3) reduces to

$$\psi_c\sim H(r)\left\{\frac{1}{i4}\int_S e^{-ik\rho\cos(\varphi-\theta)}\frac{\partial\psi}{\partial n}-\psi\frac{\partial e^{-ik\rho\cos(\varphi-\theta)}}{\partial n}\right\}ds$$
$$\equiv H(r)f(\theta,\alpha), \qquad (3')$$

where the "scattering amplitude" $f(\theta,\alpha)$ indicates the far-field response in a direction θ to a wave incident at an angle α.

We do not intend to specify the boundary conditions and seek a solution. Rather we demonstrate the relationship of the forward scattered amplitude $f(\alpha,\alpha)$ to the absorption and scattering cross sections of the cylinder.

The absorption cross section, which is defined as the time-averaged energy flow into unit length of cylinder divided by the time-averaged flux per unit area of incident wave, may be written as

$$q_a\equiv-\mathrm{Re}\frac{1}{ik}\int_0^{2\pi}\psi^*(r,\theta)\frac{\partial\psi(r,\theta)}{\partial r}rd\theta$$
$$=-\mathrm{Re}\frac{1}{ik}\int_S\psi^*(s)\frac{\partial\psi(s)}{\partial n}ds. \qquad (5)$$

The first integral is taken over a circle of radius r enclosing S; the second is taken over S. From elementary considerations we know that the integrals are equal.

Similarly the scattering cross section, which is defined as the time-averaged scattered power per unit length of cylinder divided by the time-averaged incident flux density, may be written as

$$q\equiv\mathrm{Re}\frac{1}{ik}\int_0^{2\pi}\psi_c^*(r,\theta)\frac{\partial\psi_c(r,\theta)}{\partial r}rd\theta$$
$$=\mathrm{Re}\frac{1}{ik}\int_S\psi_c^*(s)\frac{\partial\psi_c(s)}{\partial n}ds. \qquad (6)$$

It follows directly from Eqs. (5) and (6) that

$$q_a=-q-\mathrm{Re}\frac{1}{ik}\int_S\left[\psi^*\frac{\partial\psi_i}{\partial n}+\psi_i^*\frac{\partial\psi}{\partial n}\right]ds, \qquad (7)$$

where we have added the term $\psi_i^*\partial\psi_i/\partial n$ (whose integrated value vanishes) within the square brackets. From Eq. (7) we obtain the total cross section

$$Q\equiv q_a+q=-\mathrm{Re}\frac{1}{ik}\int_S\left[-\psi\frac{\partial e^{-ik\rho\cos(\varphi-\alpha)}}{\partial n}\right.$$
$$\left.+e^{-ik\rho\cos(\varphi-\alpha)}\frac{\partial\psi}{\partial n}\right]ds. \qquad (7')$$

Comparison of Eqs. (7') and (3') yields the well-known relation

$$Q=-(4/k)\,\mathrm{Re}f(\alpha,\alpha); \qquad (8)$$

thus the total cross section of the cylinder, i.e., the normalized energy flux the cylinder obtains from the incident wave and dissipates as heat and scattered radiation, is proportional to the real part of the forward scattered amplitude.[3]

If there is no absorption, then Q of Eq. (8) is to be replaced by q of Eq. (6). Letting $r\to\infty$ in the first form of q of Eq. (6) we obtain

$$\mathrm{Re}f(\alpha,\alpha)=-\frac{1}{2\pi}\int_0^{2\pi}|f(\theta,\alpha)|^2d\theta. \qquad (8')$$

We now use Eq. (8) to derive a corollary. If the cylinder is located at $x=0$, $y=y'$ rather than at the origin, we have

$$\psi_c(y')\sim H(r')e^{iky'\sin\alpha}f(\theta',\alpha),$$
$$\cos\theta'=x/r'=x/[x^2+(y-y')^2]^{\frac{1}{2}}; \qquad (9)$$

the phase factor in y' is merely the phase of the incident wave at $x=0$, $y=y'$. If $f(\theta',\alpha)$ is a slowly varying func-

[2] See Papas, reference 1, for detailed treatment of a special case.

[3] Our definition of f differs by a factor of $i/4$ from that of reference 1.

tion of y' compared to $r'-y'\sin\alpha$, then

$$\int_{-\infty}^{\infty}\psi_e(y')dy' \sim (2/k\cos\alpha)\psi_i(\theta,\alpha)f(\alpha,\alpha) \quad \text{for } x>0 \qquad (10)$$

$$\sim (2/k\cos\alpha)\psi_i(\theta,\pi-\alpha)f(\pi-\alpha,\alpha) \quad \text{for } x<0,$$

where the integral was evaluated by the method of stationary phase.[4] Consequently, in the right half-plane we have

$$2\,\mathrm{Re}\left[\psi_i^*(\theta,\alpha)\int_{-\infty}^{\infty}\psi_e(y')dy'\right] \sim \frac{4\,\mathrm{Re}[f(\alpha,\alpha)]}{k\cos\alpha} = -Q\sec\alpha, \qquad (11)$$

where the last expression follows from Eq. (8).

Equation (11) is of physical interest in connection with the transmission of a wave through a planar uniform distribution of identical parallel cylinders. If δ is the average number of scatterers in unit length of y axis and if $\delta/k\cos\alpha \ll 1$, then the transmission coefficient is approximately equal to unity plus δ times the left side of Eq. (11) (i.e., unity corresponds to the incident flux, and the left side of Eq. (11) to the interference terms of the incident and average single-scattered waves).[4] Hence by means of Eq. (11) we may write the transmission coefficient as

$$T \approx 1-Q\delta\sec\alpha, \quad \delta/k\cos\alpha \ll 1. \qquad (12)$$

This result could have been written immediately from elementary physical considerations. Q is the energy "removed" from the transmitted wave by one scatterer and $\delta\sec\alpha$ (the number of scatterers illuminated by unit area of incident wave) times Q is the energy "missing" from unit area of the transmitted wave. (We ignore here the energy scattered *into* unit area of the transmitted wave, since it is in usually small compared to Q.)

THE REFLECTION CASE

We now consider the analogous reflection problem of a plane wave incident on a surface composed of an

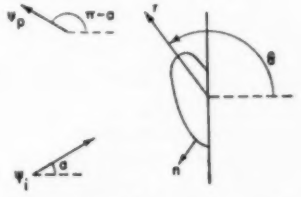


FIG. 2. Scattered reflection of a plane wave ψ_i by a smooth perfectly reflecting plane.

[4] We have shown Eq. (10) to be exact for the case of circular cylinders; see V. Twersky, "Multiple Scattering of Waves by Planar Random Distributions of Cylinders and Bosses," Mathematics Research Group, EM 58, W. S. C., New York University (1953).

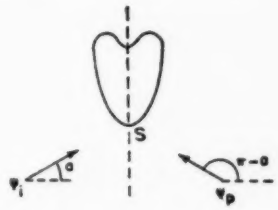


FIG. 3. The image problem: scattering of two plane waves by a cylinder.

arbitrary cylindrical protuberance on a smooth perfectly reflecting plane as in Fig. 2. We write the total field as

$$\psi(\theta,\alpha)=\psi_i+\psi_p+\psi_b, \qquad (13)$$

where ψ_i, the incident wave, is given by Eq. (2) and where

$$\psi_p=\pm\psi_i(\theta,\pi-\alpha)=\pm\exp[ik\gamma\cos(\theta-\pi+\alpha)] \qquad (14)$$

is the wave reflected from the plane in the absence of the boss; the minus sign is to be taken for $\psi=0$ at $\theta=\pi/2$ (case A), and the plus sign for $\partial\psi/\partial n=0$ at $\theta=\pi/2$ (case B). The function ψ_b, the wave scattered by the boss, satisfies the radiation condition at infinity.

For electromagnetics, case A corresponds to incident polarization perpendicular to the plane of incidence (i.e., the TM case, or "horizontal polarization"); case B corresponds to polarization parallel to the plane of incidence (i.e., the TE case, or "vertical polarization"). For acoustics, cases A and B correspond to a free and rigid plane respectively.

The problem shown in Fig. 2 is equivalent to that shown in Fig. 3 in the left-hand plane. The problem of Fig. 3 may be formulated essentially as in Sec. 2; we have

$$\psi=\psi_i+\psi_p+\psi_b=\psi_i+\psi_p-\int_S\left[G\frac{\partial\psi}{\partial n}-\psi\frac{\partial G}{\partial n}\right]ds. \qquad (15)$$

The value of Eq. (15) in the left-hand plane is identically Eq. (13).

The function ψ_b is formally identical with ψ_e of Eqs. (3) and (3′), but the present two scattering amplitudes, say $g_\pm$, differ, of course, from f of Eq. (3′) as well as from each other. The essential difference is one of symmetry. Thus, for case A, $\psi=0$ at $\theta=\pi/2$, we must have

$$g_-(\theta,\alpha)=-g_-(\pi-\theta,\alpha) \qquad (16)$$

while for case B, $\partial\psi/\partial n=0$ at $\theta=\pi/2$, we require

$$g_+(\theta,\alpha)=g_+(\pi-\theta,\alpha). \qquad (17)$$

Following the procedure of Sec. 2 we now seek a relation between the total cross section and the scattering amplitude in the direction of specular reflection of the plane, $g(\pi-\alpha,\alpha)$.

The absorption and scattering cross sections are given by Eqs. (5) and (6), except that the right-hand sides of the equations are to be divided by 2; this normalization takes into account the fact that we deal here with two incident plane waves, and facilitates extension of the final results to the problem of Fig. 2. We now have

corresponding to Eq. (7)

$$q_a=-q-\mathrm{Re}\frac{1}{2ik}\int_S\left[\psi^*\frac{\partial(\psi_i+\psi_p)}{\partial n}+(\psi_i+\psi_p)^*\frac{\partial\psi}{\partial n}\right]ds. \quad (18)$$

Hence,

$$Q=q_a+q=-\mathrm{Re}\frac{1}{2ik}\int_S\left[-\psi\frac{\partial(\psi_i+\psi_p)^*}{\partial n}+(\psi_i+\psi_p)^*\frac{\partial\psi}{\partial n}\right]ds. \quad (19)$$

For case A we have

$$\psi_i^*(s)+\psi_p^*(s)=e^{-ik\rho\cos(\varphi-\alpha)}-e^{-ik\rho\cos(\varphi-\pi+\alpha)}. \quad (20)$$

Hence, from the definition Eq. (3′) and the symmetry property Eq. (16) we obtain

$$Q_-=-(2/k)\,\mathrm{Re}[g_-(\alpha,\alpha)-g_-(\pi-\alpha,\alpha)]=(4/k)\,\mathrm{Re}[g_-(\pi-\alpha,\alpha)]. \quad (21)$$

For case B we change the sign of the second term of Eq. (20) and obtain

$$Q_+=-(2/k)\,\mathrm{Re}[g_+(\alpha,\alpha)+g_+'(\pi-\alpha,\alpha)]=-(4/k)\,\mathrm{Re}[g_+(\pi-\alpha,\alpha)]. \quad (22)$$

Equations (21) and (22) are formally identical except for the sign. Hence, for uniformity, we now define $g_-(\theta,\alpha)$ to be the negative of the function defined in Eq. (3′) (or else we may simply change the signs of ψ_i and ψ_p for case A). On suppressing the subscript $\pm$ we may write for either case

$$Q=q_a+q=-(4/k)\,\mathrm{Re}[g(\pi-\alpha,\alpha)] \quad (23)$$

which is the desired relation.

In terms of the result for a single plane wave we may rewrite Eq. (22) as

$$Q_+=-(4/k)\,\mathrm{Re}[f(\pi-\alpha,\alpha)+f(\pi-\alpha,\pi-\alpha)],$$

where the first function is evaluated at the specular angle of reflection with respect to the plane of symmetry and the second is evaluated in the forward scattered direction. The difference between the present problem and that of Sec. 2 is clear for the case of nonabsorption. In analogy to Eq. (8′) we now obtain

$$\mathrm{Re}[g(\pi-\alpha,\alpha)]=-\frac{1}{4\pi}\int_0^{2\pi}|g(\theta,\alpha)|^2d\theta=-\frac{1}{2\pi}\int_{\pi/2}^{3\pi/2}|g(\theta,\alpha)|^2d\theta. \quad (24)$$

Hence, in terms of the amplitudes corresponding to a single plane wave, we have for case B

$$\mathrm{Re}[f(\pi-\alpha,\alpha)+f(\pi-\alpha,\pi-\alpha)]=-\frac{1}{4\pi}\int_0^{2\pi}|f(\theta,\alpha)+f(\theta,\pi-\alpha)|^2d\theta. \quad (24')$$

For case A and $-\psi_i$ incident we simply change the signs of the $f(\theta,\alpha)$ terms. Hence, since $f(\alpha,\alpha)=f(\pi-\alpha,\pi-a)$, it follows from Eqs. (24′) and (8′) that‡

$$\mathrm{Re}[f(\pi-\alpha,\alpha)]=-\mathrm{Re}\frac{1}{2\pi}\int_0^{2\pi}f(\theta,\alpha)f^*(\theta,\pi-\alpha)d\theta. \quad (24'')$$

Equation (23) was derived for the problem of Fig. 3. However, it also holds for the problem of Fig. 2 by virtue of our normalization; i.e., in Fig. 2 we have only one incident wave (this eliminates the factor of $\frac{1}{2}$ we introduced in q_a of Eq. (18)), but we integrate only over $\pi/2<\theta<3\pi/2$ (this restores the factor of $\frac{1}{2}$). Hence, in using Eq. (23), we take q_a and q as in Eqs. (5) and (6) and either introduce the factor $\frac{1}{2}$ or else integrate only over the left-hand plane.

Equation (23) is a reflection theorem analogous to the "transmission theorem" of Eq. (8). The present theorem states that *the total energy cross section of an arbitrary boss on a perfectly reflecting plane is proportional to the real part of the scattering amplitude evaluated in the direction of specular reflection of the plane.*

We now use Eq. (23) to derive a corollary analogous to Eq. (11). We proceed as for Eqs. (9) and (10) with ψ_c replaced by ψ_b and f replaced by g; here only the result of Eq. (10) for $x<0$ applies. Hence, we find

$$2\,\mathrm{Re}\left[\psi_p^*\int_{-\infty}^{\infty}\psi_b(y')dy'\right]\sim\frac{4\,\mathrm{Re}[g(\pi-\alpha,\alpha)]}{k\cos\alpha}=-Q\sec\alpha \quad (25)$$

where $\psi_p=\psi_i(\theta,\pi-\alpha)$, in accordance with the convention introduced for Eq. (23). This theorem is of physical interest in connection with reflection from a uniform distribution of bosses on a plane (i.e., a "striated surface"). In analogy to the transmission coefficient of Eq. (12) we now have the reflection coefficient

$$R\approx1-Q\delta\sec\alpha,\quad \delta/k\cos\alpha\ll1, \quad (26)$$

where $Q\delta\sec\alpha$ is the energy "missing" from unit area of the wave reflected at the specular angle of the plane.[5]

The present results can be readily extended to the analogous three-dimensional problems.

‡ A generalization of Eq. (24″) has been derived by Professor M. Schiffer of Stanford.

[5] See V. Twersky, J. Appl. Phys. **24**, 659 (1953) for further discussion and specific examples.

Steady-State Solution of the Two-Dimensional Diffusion Equation for Transistors

J. S. SCHAFFNER AND J. J. SURAN
General Electric Company, Electronics Laboratory, Syracuse, New York
(Received October 15, 1953)

The diffusion equation is solved for transistors for (a) linear flow of minority carriers and (b) non-uniform flow of minority carriers due to a metal contact applied to the base region. The current transport ratio β is calculated for both of these cases as a function of frequency.

Frequency response characteristics of a transistor, as predicted by the one-dimensional diffusion equation, give excellent results for junction transistors having base length-to-base width ratios greater than ten and having negligible surface recombination rates. For transistors where these conditions are not applicable, the two-dimensional diffusion equation must be used in the calculation of β.

INTRODUCTION

IT has been shown that the process by which a signal is transferred from the emitter to the collector of a transistor is, essentially, one of diffusion.[1] This is to say that minority carriers (electrons in $n-p-n$ transistors and holes in $p-n-p$ types) move through the base region of a transistor by means of a concentration gradient rather than an electric field gradient.

The diffusion equation which applies to transistors may be derived from the continuity equation and the current-flow equation. For the $n-p-n$ transistor, the continuity of flow equation is[1]

$$q\frac{\partial m}{\partial t}=\text{div}\mathbf{i}+q\frac{m_e-m}{\tau_m}, \tag{1}$$

where m is the minority carrier concentration (L^{-3}), $\mathbf{i}$ is the current density, τ_m is the carrier lifetime, m_e is the thermal equilibrium concentration of the minority carriers, and q the electronic charge. The current-flow equation is given by

$$\mathbf{i}=qD_m\,\text{grad}m, \tag{2}$$

where D_m is the carrier diffusion constant. Combining Eqs. (1) and (2) gives

$$\frac{\partial m}{\partial t}=D_m\nabla^2 m+\frac{m_e-m}{\tau_m}. \tag{3}$$

Equation (3) is the general diffusion equation.

SOLUTION OF THE ONE DIMENSIONAL DIFFUSION EQUATION

Figure 1 illustrates the physical construction of a junction transistor. In most practical cases, the width of the base region W will be very small compared to the length B of the base. Therefore, it may be assumed that the gradient of the minority carrier density is essentially perpendicular to the junction surfaces.

Written for only one space dimension, Eq. (3) becomes

$$\frac{\partial m}{\partial t}=D_m\frac{\partial^2 m}{\partial x^2}+\frac{m_e-m}{\tau_m}. \tag{4}$$

There is ample experimental justification for assuming that the one-dimensional diffusion equation, given by (4), will lead to a good first approximation for the behavior of junction transistors. Equation (4) may be solved on a steady-state basis for a minority carrier concentration consisting of a dc and an ac component

$$m=M_0(x)+M(x)e^{j\omega t}, \tag{5}$$

where $M_0(x)$ is the dc part of the carrier concentration and $M(x)\cdot e^{j\omega t}$ the ac part. Substituting Eq. (5) into (4) results in

$$j\omega Me^{j\omega t}=D_m\left[\frac{\partial^2 M_0}{\partial x^2}+\frac{\partial^2 M}{\partial x^2}e^{j\omega t}\right]+\frac{m_e-M_0-Me^{j\omega t}}{\tau_m}. \tag{6}$$

The ac and the dc equations separate out as

$$\left\{0=D_m\frac{\partial^2 M_0}{\partial x^2}+\frac{m_e-M_0}{\tau_m}\right\}\text{dc} \tag{7}$$

$$\left\{0=D_m\frac{\partial^2 M}{\partial x^2}-\left(j\omega+\frac{1}{\tau_m}\right)M\right\}\text{ac}. \tag{8}$$

The solution of (7) for M_0 will give the carrier concentration in the base region of a transistor due to the dc component of emitter current. Equation (8) will give the ac component of the minority carrier concentration due to the ac component of emitter current.

Before Eq. (8) can be solved for a specific case, it is necessary to know the boundary conditions which apply to the physical problem under consideration. For the one-dimensional case given by Eq. (8), it is

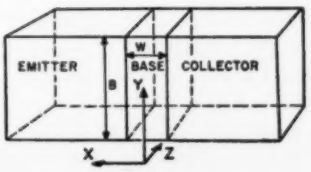


FIG. 1. Construction of junction transistor.

[1] W. Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950).

generally assumed[1] that the following boundary conditions apply:

(a) The carrier concentration at the collector boundary is zero, and therefore $M/\text{collector}=0$.

(b) The gradient of the minority carrier density at the emitter base boundary is known. The current density $\mathbf{i}$ corresponding to this gradient is

$$\mathbf{i}=-qD_m\frac{\partial m}{\partial x}. \tag{9}$$

The boundary condition (b) is, if $Je^{j\omega t}$ is the ac component of $\mathbf{i}$,

$$\left.\frac{\partial M}{\partial x}\right|_{\text{emitter}}=\frac{-J}{qD_m}.$$

Equation (8) may be solved in the usual manner for these boundary conditions. The complete solution is

$$M(x)=\frac{-JL_m}{qD_m\cosh\frac{W}{L_m}\lambda}\cdot\sinh\frac{x}{L_m}\lambda \tag{10}$$

where $\lambda=(1+j\omega\tau_m)^{\frac{1}{2}}$, $L_m^2=\tau_mD_m$, W is the base width, and x is a space coordinate, the origin of which is taken at the collector (Fig. 1). To find the current transport ratio β, the ratio of collector current to emitter current must be taken. From Eq. (10), β is found to be

$$\beta=\frac{\left.\frac{\partial m}{\partial x}\right|_{x=0}}{\left.\frac{\partial m}{\partial x}\right|_{x=W}}=\operatorname{sech}\frac{W}{L_m}(1+j\omega\tau_m)^{\frac{1}{2}}. \tag{11}$$

The current transport ratio β is, for many transistors, equal to the short circuit current amplification α.

SOLUTION OF THE TWO-DIMENSIONAL DIFFUSION EQUATION

The two-dimensional diffusion equation is

$$\frac{\partial m}{\partial t}=D_m\left[\frac{\partial^2 m}{\partial x^2}+\frac{\partial^2 m}{\partial y^2}\right]+\frac{m_e-m}{\tau_m}. \tag{12}$$

Equation (12) differs from the one-dimensional equation in that the two-dimensional representation recognizes that minority carriers may diffuse in two directions, i.e., that carrier current may flow into the base electrode as well as into the collector. [Equation (12) implies that the z component of minority carrier flow, everywhere in the base region, is zero, i.e., $\partial m/\partial z\equiv 0$ and therefore $\partial^2 m/\partial^2 z\equiv 0$.] Rigorously, Eq. (12) should be used to solve for the transistor transit-time effects when diffusion into the base electrode must be accounted for. Furthermore, if it is desired to know what effect the dimensioning of the base region in a transistor has upon the transistor parameters (particularly β), resort must be made to Eq. (12).

Equation (12) is a partial differential equation in two dimensions which may be solved, in the steady state, by the same general technique which was employed for the solution of the one-dimensional equation. Thus, an emitter current composed of a dc and an ac component may be assumed to result in a carrier concentration which is the linear superposition of a time invariant and a time-variant function, as given by Eq. (5). In the two-dimensional case, both M_0 and M will be functions of the two space dimensions, x and y. Substitution of Eq. (5) into Eq. (12) leads to the dc and ac differential equations given below, which correspond to Eqs. (7) and (8) for the one-dimensional problem.

$$\left\{\frac{\partial^2 M_0}{\partial x^2}+\frac{\partial^2 M_0}{\partial y^2}+\frac{m_e-M_0}{\tau_mD_m}=0\right\}\text{dc} \tag{13}$$

$$\left\{\frac{\partial^2 M}{\partial x^2}+\frac{\partial^2 M}{\partial y^2}-\left(\frac{1}{\tau_mD_m}+\frac{j\omega\tau_m}{\tau_mD_m}\right)M=0\right\}\text{ac}. \tag{14}$$

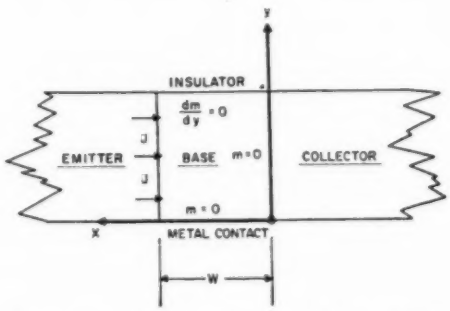


FIG. 2. Two-dimensional geometry for the triode.

Equation (14) must be solved in order to determine the variation of β with frequency.

Partial differential equations of the type given by Eq. (14) may be readily solved if it is assumed that the space function $M(x,y)$ consists of a product of two separate space functions $X(x)$ and $Y(y)$. Thus,

$$M(x,y)=X(x)\cdot Y(y). \tag{15}$$

Substitution of Eq. (15) into Eq. (14) results in

$$Y(y)\frac{\partial^2 X}{\partial x^2}+X(x)\frac{\partial^2 Y}{\partial y^2}-\frac{\lambda^2}{L_m^2}X(x)Y(y)=0, \tag{16}$$

[where $L_m^2=\tau_mD_m$ and $\lambda=(1+j\omega\tau_m)^{\frac{1}{2}}$]. If Eq. (15) is divided by $X(x)Y(y)$, the following equation is obtained

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2}+\frac{1}{Y}\frac{\partial^2 Y}{\partial y^2}-\frac{\lambda^2}{L_m^2}=0. \tag{17}$$

Equation (17) may be solved by separation of variables.

Hence,

$$\frac{1}{X}\frac{\partial^2 X}{\partial x^2}-\frac{\lambda^2}{L_m{}^2}=K^2 \tag{18}$$

$$\frac{1}{Y}\frac{\partial^2 Y}{\partial y^2}=-K^2, \tag{19}$$

where K is an eigenvalue constant. The space functions X and Y are the solutions of the equations

$$\frac{\partial^2 X}{\partial x^2}-\left(K^2+\frac{\lambda^2}{L_m{}^2}\right)X=0, \tag{20a}$$

$$\frac{\partial^2 Y}{\partial y^2}+K^2 Y=0. \tag{20b}$$

In order to solve Eqs. (20a) and (20b) for a particular case, four boundary conditions are required. Figure 2 shows the base region of a transistor. The boundary conditions at the emitter and collector are assumed to be the same as for the one-dimentional case, i.e., that the

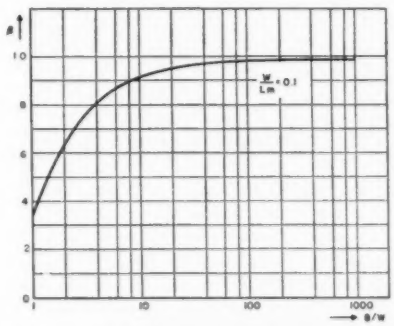


FIG. 3. Variation of low-frequency β with base dimensions.

time-varying component of emitter current density is a known periodic time function, uniform across the emitter junction, and that the ac minority carrier concentration at the collector is zero. It is further assumed that the ac component of the minority carrier concentration at the metal-semiconductor boundary of the base region is zero, and that there is no y component of minority carrier current at the insulator semiconductor boundary of the base region. The total minority carrier current (per unit length in the z direction) which flows into the collector will be the integrated current density in the x direction at the base-collector boundary. Thus

$$I_c'=-qD_n\int_0^{y=B}\frac{\partial M}{\partial x}(x,y)\bigg|_{x=0}dy. \tag{21}$$

Similarly, the minority carrier current flowing into the base lead will be given by:

$$I_b'=-qD_n\int_0^{x=W}\frac{\partial M}{\partial y}(x,y)\bigg|_{y=0}dx \tag{22}$$

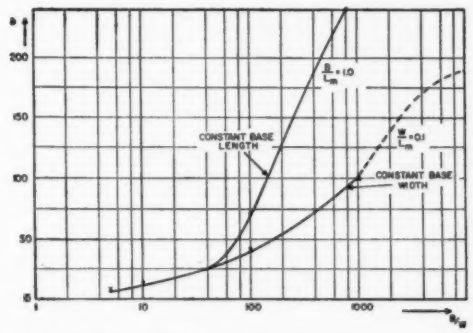


FIG. 4. Variation of low-frequency b with base dimensions.

Equations (20a) and (20b), subject to the boundary conditions stated above, are solved for $M(x,y)$ in Appendix II. The current transport ratio β is found to be

$$\beta=\frac{I_c'}{I_e'}=\frac{8}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}\operatorname{sech} r\left(1+\left(\frac{n\pi}{2}\frac{s}{r}\right)^2+j\omega\tau_m\right)^{\frac{1}{2}}$$

$$n=1,3,5,7\cdots \tag{23}$$

where

$$r=\frac{W}{L_m},$$

$$s=\frac{W}{B}.$$

For very small ratios of base width to base lengths, $s\ll 1$ and Eq. (23) becomes identical to the β derived from the one-dimensional diffusion equation [Eq. (11)].

Using Eq. (23) it is interesting to plot the variation of β against the ratio of base length to base width $(1/s)$. This relationship is illustrated for low frequencies $(\omega\to 0)$, and for $r=0.1$ in Fig. 3. The current-amplification for a grounded-emitter stage is approximately given by

$$b=\frac{\beta}{1-\beta}. \tag{24}$$

Figure 4 shows b plotted against $1/s$ first for the condition that the width (W) is fixed and $r=0.1$ and second, for the condition that the length (B) is fixed and $B/L_m=1.0$.

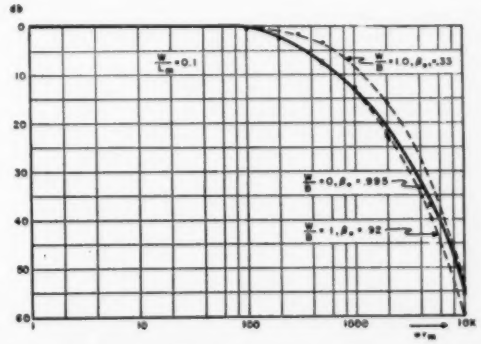


FIG. 5. Variation of β with frequency-normalized scale.

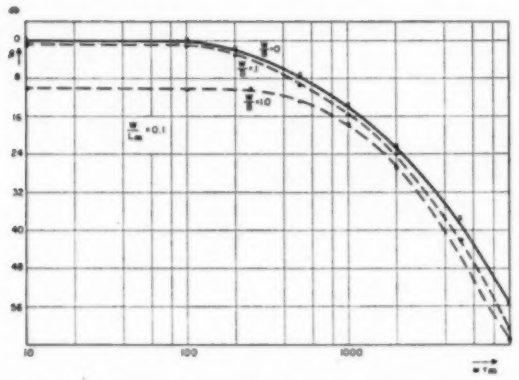


FIG. 6. Variation of β with frequency-relative scale.

In Fig. 5 the variation of β with frequency is plotted as predicted by Eq (23). The solid line ($s=0$) in Fig. 5 is the function, $\beta(\omega)$, which is predicted by the one-dimensional diffusion equation. The dotted line, for $s=0.1$, is seen to be in very close agreement with the $s=0$ curve, which verifies the assumption that the one-dimensional diffusion equation is an adequate approximation to the carrier diffusion phenomena in present-day junction transistors. The symbol β_0 will be used in the following for the low-frequency value of β.

For low β_0 transistors Eq. (23) predicts a greater bandwidth (of the $\beta(\omega)$ curve) than for high β_0 transistors. This is illustrated in Fig. 5, and agrees with experimental observations. From this apparent increase in bandwidth it may not be deduced, however, that a low-β_0 transistor can be used as an oscillator at higher frequencies than those obtainable with a high-β_0 transistor. The curves in Fig. 5 are redrawn, on a relative basis, in Fig. 6. No crossing of the lines for high-β_0 by the curves for low-β_0 is observed, and so, despite the apparent increase in bandwidth of the low β_0 units, it is obvious that the greater bandwidth of low-β_0 transistors cannot be utilized practically for higher-frequency amplification purposes.

Figure 7 illustrates both the phase-shift and attenuation curves for β and b [b is given by Eq. (24)] as predicted by the two-dimensional diffusion equation. In the range shown these results are very close to the curves which would be obtained with the one-dimensional diffusion equation (within 5 percent).

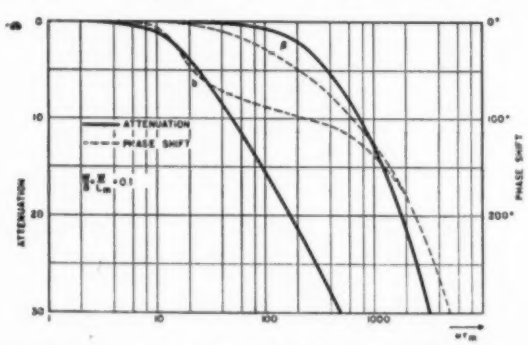


FIG. 7. Attenuation and phase shift of β and b with frequency.

CONCLUSION

The current transfer function, or β, of junction transistors has been derived theoretically from the one-dimensional diffusion equation and has checked experimental measurements very closely. However, the theory based upon the one-dimensional solution begins to deviate from experiment when base current must be taken into account. Furthermore, the one-dimensional equation does not account for the variation of current amplification with the base dimensioning. In order to thoroughly investigate the semiconductor triode, the two-dimensional diffusion equation must be solved.

ACKNOWLEDGMENT

The authors wish to thank Dr. Paul Weiss and Dr. J. S. Kerr, of the Electronics Laboratory, for their helpful suggestions.

This work has been supported by the U. S. Air Force Air Research and Development Command, U. S. Air Force Air Materiel Command, Army Signal Corps and Navy Bureau of Ships under contract AF 33 (600)-17793.

APPENDIX I. DIMENSIONS

Table I gives the symbols and the dimensional exponents of the physical parameters used in this article.

TABLE I.

Parameter	Symbols	L	T	q
Carrier concentration	m,M	-3	0	0
Diffusion constant	D_m	2	-1	0
Carrier lifetime	τ_m	0	1	0
Carrier diffusion length	L_m	1	0	0
Base width	W	1	0	0
Base length	B	1	0	0
Current density	$\mathbf{i},J$	-2	-1	1
Current per unit length	I'	-1	-1	1

APPENDIX II. SOLUTION OF THE TWO-DIMENSIONAL DIFFUSION EQUATION

The time-dependent two-dimensional diffusion equation, given by Eq. 14, was separated into two linear differential equations [Eqs. (20a) and (20b)] by a separation-of-variables technique. Both Eqs. (20a) and (20b) must satisfy the boundary conditions (see Fig. 2), which are algebraically stated as follows:

$$M(0,y)=0, \tag{IIa}$$

$$M(x,0)=0, \tag{IIb}$$

$$-qD_m\frac{\partial M}{\partial x}(W,y)=J, \tag{IIc}$$

$$\frac{\partial M}{\partial y}(x,B)=0. \tag{IId}$$

The general solution of Eq. (20a) is

$$X=C_1\exp(K^2+\lambda^2/L_m^2)^{\frac{1}{2}}x + C_2\exp-(K^2+\lambda^2/L_m^2)^{\frac{1}{2}}x. \quad \text{(IIe)}$$

Similarly, the general solution of Eq. (20b) is

$$Y=C_3e^{jKy}+C_4e^{-jKy}. \quad \text{(IIf)}$$

The general solution of Eq. (14) is, therefore

$$M=[C_1\exp(K^2+\lambda^2/L_m)^{\frac{1}{2}}x+C_2\exp-(K^2+\lambda^2/L_m^2)^{\frac{1}{2}}x] \times(C_3e^{jKy}+C_4e^{-jKy}). \quad \text{(IIg)}$$

Applying the boundary conditions (IIa), (IIb), and (IId) gives

$$M(x,y)=C_n\sin\frac{n\pi}{2B}y\sinh\left(\left(\frac{n\pi}{2B}\right)^2+\frac{\lambda^2}{L_m^2}\right)^{\frac{1}{2}}x, \quad \text{(IIh)}$$

where $n=1, 3, 5, 7, \cdots$. The complete solution is therefore

$$M(x,y)=\sum_{n=1}^{\infty}C_{(2n-1)}\sin\frac{(2n-1)\pi}{2B}y \times\sinh\left(\left(\frac{(2n-1)\pi}{2B}\right)^2+\frac{\lambda^2}{L_m^2}\right)^{\frac{1}{2}}x. \quad \text{(IIi)}$$

The constants C_n must now be determined from the remaining boundary condition (IIc).

From Eq. (IIi) we obtain for $x=W$

$$\frac{\partial M}{\partial X}=\sum_{n=1}^{\infty}C_{2n-1}\sin\frac{(2n-1)\pi}{2B}yK_1\cosh K_1W, \quad \text{(IIj)}$$

where

$$K_1=\left(\left(\frac{(2n-1)\pi}{2B}\right)^2+(\lambda/L_m)^2\right)^{\frac{1}{2}}. \quad \text{(IIk)}$$

A function that has the value $+1$ in the range $0<\phi<\pi$ (and -1 in the range $\pi<\phi<2\pi$) is

$$f(\phi)=\sum_{n=1}^{\infty}\frac{2}{(2n-1)}\sin(2n-1)\phi. \quad \text{(IIl)}$$

The value of $\partial M/\partial x$ should be $-J/qD_m$ in the range $0<y<B$ (condition IIc). A comparison between equations (IIj) and (IIl) shows that the values of the coefficients C_n are

$$C_{(2n-1)}=\frac{4J}{qD_m\pi(2n-1)}\frac{\operatorname{sech}K_1W}{K_1}. \quad \text{(IIm)}$$

Combining Eq. (IIj) and (IIm), the complete solution for the ac component of the minority carrier concentration is

$$M(x,t)=\frac{-4J}{qD_m\pi}\sum_{n=1}^{\infty}\frac{\sin\frac{(2n-1)\pi}{2B}y\sinh K_1X}{(2n-1)K_1\cosh K_1W}. \quad \text{(IIn)}$$

The collector current can be calculated from Eq. (11). The ac component of the current density at the collector is

$$i_c=\frac{4J}{\pi}\sum_{n=1}^{\infty}\frac{\sin\frac{(2n-1)\pi}{2B}y}{(2n-1)\cosh K_1W}. \quad \text{(IIo)}$$

The total collector current (per unit length in the Z direction) is [Eq. (11)]

$$I_c'(t)=\frac{8BJ}{\pi^2}e^{j\omega t}\sum_{n=1}^{\infty}\frac{\operatorname{sech}K_1W}{(2n-1)^2}. \quad \text{(IIp)}$$

Since the total current (per unit length) at the emitter boundary is (boundary condition IIc)

$$I_e'(t)=JBe^{j\omega t} \quad \text{(IIq)}$$

the ratio of collector current to emitter current is

$$\beta=\frac{8}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{(2n-1)^2} \times\operatorname{sech}\frac{W}{L_m}\left(1+\left(\frac{(2n-1)\pi L_m}{2B}\right)^2+j\omega\tau_m\right)^{\frac{1}{2}} \quad \text{(IIr)}$$

using the symbols $r=W/L_m$ and $s=W/B$ this equation is

$$\beta=\frac{8}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{(2n-1)^2} \times\operatorname{sech}r\left[1+\left(\frac{(2n-1)\pi s}{2r}\right)^2+j\omega\tau_m\right]^{\frac{1}{2}}. \quad \text{(IIs)}$$

BIBLIOGRAPHY

W. Shockley, Bell System Tech. J. **28**, 435, (1949).
Shockley, Sparks, and Teal, Phys. Rev. **83**, 151 (1951).
E. L. Steele, Proc. Inst. Radio Engrs. **40**, 1424 (1952).

Diffusion Coefficient, Solubility, and Permeability for Helium in Glass

W. A. ROGERS, R. S. BURITZ, AND D. ALPERT
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received October 23, 1953)

A method is described for simultaneously measuring the diffusion coefficient, solubility, and permeability for a given gas in a single sample of solid material. The method utilizes solutions of the diffusion equation (for both plane and cylindrical geometry) which make it easily possible to calculate the diffusion parameters from data taken either in the period of steady-state gas flow or in the early part of the transient period. Using ultra-high vacuum techniques, it is possible to make the measurements with extremely small amounts of gas and without errors due to leaks or contamination in the gas handling system.

Values for the diffusion parameters have been determined for helium in Pyrex glass (Corning 7740), and comparisons with previous measurements are made. Directly measured values of the diffusion coefficient are given for the first time. The diffusion coefficient and permeability both exhibit a temperature dependence of the form $\exp(-E/kT)$ both having the same activation energy (6.4×10^3 cal/mole) within the experimental error. The solubility exhibits little or no variation with temperature over the range 25°C to 300°C, and is in good agreement with the only previous measurement, carried out at 515°C.

I. INTRODUCTION

THE permeation of glass by various gases has been measured experimentally over the past fifty years for a wide variety of glasses. In 1924, C. C. van Voorhis[1] reported a study in which attention was centered on the relation of permeability to chemical composition of the glass. His work appears to be the beginning of a series of similar investigations[2] aimed at an understanding of the fundamental physical and chemical factors influencing permeation. In the same year, Williams and Ferguson[3] followed up their own permeability measurements of two years earlier with measurements of the solubility of helium in silica and Pyrex,[4] pointing out the importance of understanding the relation between permeability, solubility, and diffusion coefficients. Despite this early recognition of the underlying principles, the numerous investigations of permeability during the past three decades have included no independent measurements of solubility or diffusion coefficient. The single value of solubility obtained by Williams and Ferguson at 515°C has remained the only available one for helium in Pyrex. Values of the diffusion coefficient can be calculated by dividing permeability values of other investigators by the one available solubility value. However, a survey of the literature on permeability reveals that in the case of helium in Pyrex the values reported by various observers differ by as much as a factor of twenty.

Recent experiments in this laboratory have indicated the need for better information concerning the diffusion of gas through glass. In typical ultra-high vacuum systems sealed and evacuated to pressures of the order of 10^{-10} mm Hg,[5] we have observed rates of pressure rise of the order of 10^{-11} mm Hg per minute. These observed rates have suggested that they might be due to the permeation of atmospheric helium through the glass walls of the vacuum system. A simple test of this hypothesis was made by placing a sealed, evacuated ionization gauge in a larger vessel and then evacuating the larger vessel with a mechanical pump.[6] The rise rate in the ionization gauge, which had been constant for 75 days previously, gradually decreased until after 60 days it was no longer observable. The total disappearance of the rise rate verified that it was due to gas entering from outside, and the long time delay indicated that the leakage occurred by a true diffusion rather than by effusion through a definite hole. By analysis of the data we obtained values of the diffusion coefficient and solubility for the unknown atmospheric gas in the glass (Corning 7740) of which the ionization gauge was made and should have liked to compare them with known values for helium in the same type of glass. Because there appeared to be no reliable values of these quantities, the present set of experiments was proposed in which we would measure simultaneously the diffusion coefficient, solubility, and permeability for helium in Pyrex using ultra-high vacuum techniques. The low residual pressure and very small rise rates which can be obtained with the ultra-high vacuum techniques assure the absence of leaks and other sources of contamination and make possible the detection of very small quantities of gas. This permits the utilization of data obtained during the initial transient period of gas diffusion.

II. PRINCIPLE OF THE EXPERIMENT

For the measurement of permeability one usually has an arrangement such as that shown in Fig. 1 with a membrane made of the solid to be studied separating the

[1] C. C. van Voorhis, Phys. Rev. **23**, 557 (1924).
[2] W. D. Urry, J. Am. Chem. Soc. **54**, 3887 (1932); Lord Rayleigh, Proc. Roy. Soc. (London) **156A**, 350 (1936); N. W. Taylor and W. J. Rast, J. Chem. Phys. **6**, 612 (1938); G. P. Baxter, J. Am. Chem. Soc. **61**, 1597 (1939); P. L. Smith and N. W. Taylor, J. Am. Ceram. Soc. **23**, 139 (1940); T. F. Newkirk and F. V. Tooley, J. Am. Ceram. Soc. **32**, 272(1949); F. J. Norton, J. Am. Ceram. Soc. **36**, 90 (1953).
[3] G. A. Williams and J. B. Ferguson, J. Am. Chem. Soc. **44**, 2160 (1922); J. Am. Chem. Soc. **46**, 635 (1924).
[4] Where the word Pyrex appears without further clarification in the earlier literature, we assume that it refers to Pyrex-brand (Corning 7740), which is the type used in the present experiments.
[5] D. Alpert, J. Appl. Phys. **24**, 860 (1953).
[6] D. Alpert and R. S. Buritz, J. Appl. Phys. **25**, 202 (1954).

two chambers of a vacuum system. The gas to be studied is introduced into one of these chambers, and after waiting long enough to establish a constant rate of flow through the membrane, the steady rate of accumulation of gas in the other chamber is measured with a suitable vacuum gauge. The current density j of the gas diffusing through the membrane is related to the pressure readings in the volume V in which the diffused gas accumulates, by

$$j=\frac{V}{A}\frac{dp}{dt}, \tag{1}$$

where A is the area of the membrane. In a plane membrane, the current density in the steady state is assumed to depend on the gas pressure p_1 and the glass thickness d through the relation

$$j=kp_1/d, \tag{2}$$

provided the pressure p in V is always much smaller than p_1. The constant of proportionality k is called the permeability. Equation (2) is a special case of the fundamental diffusion law

$$j=-D\nabla n, \tag{3}$$

which is sometimes referred to in the literature as Fick's law. In Eq. (3) n represents the concentration of gas inside the glass membrane. The constant D, defined by this equation, is the diffusion coefficient.

For the present experiments, in which the current density and concentration vary with the time, it is necessary to obtain a solution of the time dependent diffusion equation

$$D\nabla^2 n=\frac{\partial n}{\partial t}, \tag{4}$$

which can be derived from Eq. (3) using the particle continuity equation.[7] The particular solution of Eq. (4) which is applicable to these experiments is specified by the conditions that the entire apparatus is initially devoid of gas; at time zero the gas is introduced suddenly at a pressure p_1 into the volume V_1 as indicated in Fig. 1, and is kept at constant pressure as long as the experiment continues. The pressure p in V remains very small in comparison with p_1; consequently the boundary value of n at this surface may be taken as identically zero.

To relate the pressure p_1 to the boundary values of n at the surface of V_1, we define the solubility of the gas in glass as the ratio of the concentration of gas particles in glass to the external gas pressure when these two phases are in static equilibrium. Thus, a piece of glass exposed to gas under a pressure p_1 will absorb or dissolve the gas until the concentration of dissolved gas reaches a saturation value n_1 given by

$$n_1=Sp_1, \tag{5}$$

where S is the solubility. The concentration n of dissolved gas is expressed as the pressure of the same concentration of gas in the free state at the standard temperature (0°C).[8] In the case where diffusion through the glass is taking place, the concentration n does not achieve a uniform value of n_1 throughout the glass. However, since the transfer of gas particles between the surface layer and the free state in the adjacent volume V_1 is very rapid in comparison with the diffusion in the interior of the glass, it is reasonable to assume, as is done in this paper, that the equilibrium value of n_1 is immediately established at the surface.

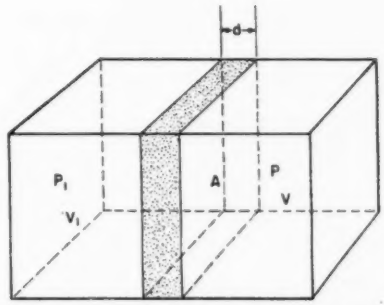


FIG. 1. Schematic diagram of apparatus for diffusion measurement in plane geometry.

A. Solutions in Plane Geometry

In the case of a plane glass wall of thickness, d, Eq. (4) can be solved by the commonly used method of assuming a solution in the form of Fourier series in the position variable with exponential time factors[9] and assigning coefficients to meet the boundary requirements. The solution of Eq. (4) obtained in this manner when substituted into Eq. (3) leads to the result

$$\frac{dp}{dt}=\frac{ADS}{Vd}p_1\{1+\sum_{m=1}^{\infty}2\cos\pi m\exp(-m^2\pi^2Dt/d^2)\}. \tag{6}$$

By inspection of Eq. (6) it can be seen that after a sufficiently long time dp/dt attains a constant value of

$$\frac{dp}{dt}=\frac{A}{V}DS\frac{p_1}{d}. \tag{7}$$

The product DS is identical with the permeability k as can be seen by comparing with Eq. (2). Substituting k for DS in Eq. (7) one obtains

$$k=\frac{Vd}{A}\frac{1}{p_1}\frac{dp}{dt}, \tag{8}$$

in which it is evident that the dimensions of k are length squared over time, *viz.*, cm²/sec.

[7] Discussion of Eq. (4) and its relation to Eq. (3) are given in numerous standard texts on heat or fluid flow. See, for example, Carslaw and Jaeger, *Conduction of Heat in Solids* (Clarendon Press, Oxford, 1947).

[8] See Appendix for a discussion of units.

[9] E.g., Carslaw and Jaeger, see reference 7, Chapter III.

By integrating Eq. (6) from time zero to time t, one obtains for the pressure p at time t,[10]

$$p=\frac{ADS}{Vd}p_1\left\{t-\frac{d^2}{6D}-\frac{2d^2}{\pi^2 D}\times\sum_{m=1}^{\infty}(-1)^m\frac{\exp(-m^2\pi^2Dt/d^2)}{m^2}\right\}. \quad (9)$$

Again it is evident that after sufficient time has elapsed the exponential terms all become negligibly small and the pressure becomes linear with time. Then Eq. (9) reduces to

$$p=\frac{ADS}{Vd}p_1(t-t_c), \quad (10)$$

where $t_c=d^2/6D$. Equation (10) is referred to in what follows as the late (i.e., large t) approximation. The characteristic time t_c is a measure of the time required to establish a constant flow.

If we plot experimental values of pressure against time over an interval sufficiently long to identify the straight line of Eq. (10), the intercept t_c of this line on the time axis allows a determination of the diffusion coefficient from the simple relation

$$D=d^2/6t_c. \quad (11)$$

Similarly the intercept p_c on the pressure axis affords a simple calculation of the solubility from

$$S=-\frac{6V}{Ad}\frac{p_c}{p_1}. \quad (12)$$

It is evident from Eq. (12) that the solubility is a dimensionless number, while in Eq. (11) one sees that the dimensions of D are, like those of k, cm²/sec.

In the above results we have one simple and straightforward way of getting D, S, and k. It is only necessary to admit gas into V_1, record the pressure in V as a function of time until a straight line is determined, then extrapolate this line graphically to locate the characteristic time t_c and the characteristic pressure p_c. From these intercepts D and S are calculated by means of Eqs. (11) and (12). From the slope of the line a value of k is found according to Eq. (8).

In spite of the resulting simplicity of computation, the method as described so far has at least one significant disadvantage: at room temperature the characteristic time for diffusion of helium through one millimeter thickness of Pyrex glass is approximately one week; the time required to establish the linear relation between the pressure and the time is several weeks. Not only is this inconveniently long, but also it is difficult to assure stability of the glass temperature and of the electronic equipment associated with the ionization gauge over such long periods.

To avoid these difficulties an alternative treatment of the exact solution in Fourier series [Eq. (6)] was suggested by T. Holstein in which a transformation formula[11] was applied to the right side of this equation leading to the form

$$\frac{dp}{dt}=\frac{2A}{V}Sp_1(D/\pi t)^{\frac{1}{2}}\sum_{m=0}^{\infty}\exp[-(d^2/4Dt)(2m+1)^2]. \quad (13)$$

Because of the inverted placement of t in the exponentials, this series converges most rapidly for very small values of t rather than for large values as in Eqs. (6) and (9). For times sufficiently short, Eq. (13) may be approximated by neglecting all but the leading term in the series of exponentials. Then it is convenient to multiply by $t^{\frac{1}{2}}$, and take the logarithms on both sides. Thus,

$$\ln t^{\frac{1}{2}}\frac{dp}{dt}=\ln[(2A/V)Sp_1(D/\pi)^{\frac{1}{2}}]-d^2/4Dt. \quad (14)$$

Equation (14) will be referred to in the following as the early approximation.

By plotting the quantity on the left side of Eq. (14) against reciprocal time, a straight line is determined. From the slope of this line, D can be obtained according to the relation

$$\text{slope}=-d^2/4D. \quad (15)$$

After D has been determined from Eq. (15) the solubility S can be calculated by solving Eq. (14) for S.

$$S=\left(\frac{\pi}{D}\right)^{\frac{1}{2}}\frac{V}{2A}\frac{1}{p_1}\left(t^{\frac{1}{2}}\frac{dp}{dt}\right)\exp(d^2/4Dt) \quad (16)$$

The distinction between early and late times as applied to the two foregoing approximations is clarified by comparing the value of the ordinate in each type of graph obtained from the approximation with that obtained from the corresponding exact solution. At $t=2.7t_c$ the error in either approximation becomes 1.2 percent. This particular instant may be regarded as the juncture of the validity ranges of the two approximate forms.

B. Solutions in Cylindrical Geometry

Although the results obtained in Sec. A above cover satisfactorily the whole range of time during which data can be taken in the case of plane geometry, there remains some doubt as to their accuracy when applied to the cylindrical geometry which is more easily handled

[10] The term $(d^2/6D)$ results from the summation of the series $\sum_{m=1}^{\infty}(-1)^m/m^2$ which appears in the evaluation of the integrated expression at the lower limit of the integration. In Westinghouse Research Report 411-9-D, T. Holstein has pointed out that this series may be regarded as the Fourier series representation of the function $(\pi^2/4)[(1-x)^2-\frac{1}{3}]$, $0<x\leq 1$, evaluated at $x=1$.

[11] The formula is given in Courant-Hilbert, *Methoden der Mathematischen Physik* (Julius Springer, Berlin, 1931), second edition, pp. 63, 64; and discussed in this particular application by T. Holstein, reference 10.

in the laboratory. Superficially, this case seems much more difficult analytically because it leads to rather complicated expressions involving Bessel's functions when treated by the Fourier series method. However, operational methods based on the use of the Laplace transformation[12] provide a means of obtaining approximate solutions of the time dependent diffusion equation [Eq. (4)] corresponding to the above early and late approximations without actually handling the infinite series in detail. Only a limited description of the asymptotic behavior of the Bessel function is required for the treatment of the cylindrical case.

Figure 2 illustrates the experimental situation in cylindrical geometry which is equivalent to that of the plane geometry of Fig. 1. The plane membrane of area A, and thickness d, is replaced by the cylindrical membrane of height h, outer and inner radii, a and b, respectively, and thickness $d=a-b$. The surfaces at a and b are coaxial. For radial diffusion of gas in either direction the approximation which is valid at late times is

$$p=\frac{2\pi hDS}{V\ln R}p_1\left[t-\frac{d^2}{6D}Z(R)\right], \tag{17}$$

where R is the ratio of a to b, and the function $Z(R)$ is[13]

$$Z(R)=(3/2)[(R^2+1)\ln R+1-R^2]/(R-1)^2\ln R. \tag{18}$$

Expansion of the numerator and denominator of $Z(R)$ separately in Taylor's series at $R=1$ reveals that $Z(1)=1$ in agreement with the result expressed in Eq. (10) for plane geometry (infinite radius of curvature, finite thickness).

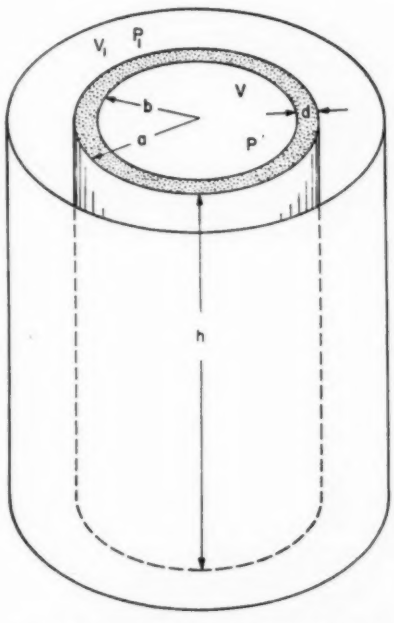


FIG. 2. Schematic diagram of apparatus for diffusion measurement in cylindrical geometry.

A very significant feature of $Z(R)$ is its remarkably slow variation with R. Taking the values of R actually used in our apparatus as typical examples, we find that for one-in. tubing with $\frac{1}{16}$-in. wall thickness $R=1.14$ and $Z(R)=1.0003$; for $\frac{1}{4}$-in. capillary of 1-mm bore, $R=6.4$ and $Z(R)=1.048$. Inspection of Eq. (18) reveals that as R becomes very large, $Z(R)$ approaches but never exceeds $\frac{3}{2}$.

In the early approximation, the solution in cylindrical geometry is identical to that in plane geometry [Eq. (14)] provided the area A is evaluated for the geometric mean radius $(ab)^{\frac{1}{2}}$.

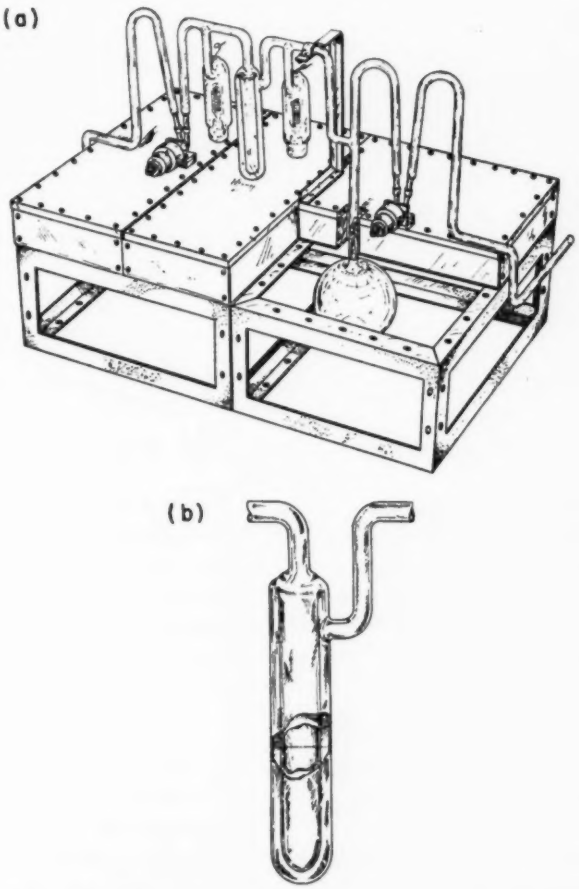


FIG. 3. (a) Sketch of experimental vacuum system used in diffusion measurements. (b) Detail of diffusion cell.

III. EXPERIMENTAL ARRANGEMENTS

The vacuum system in which the measurements were made is shown in Fig. 3(a). It consisted of two separate volumes having in common the glass wall through which the measured diffusion took place. Each of these volumes had its own vacuum pump, ion gauge,[14] and high-vacuum valve.[15] The measured diffusion took place through the

[12] See for example: Carslaw and Jaeger, reference 7 or *Operational Methods in Applied Mathematics* (Oxford University Press, London, 1941); also M. F. Gardner and J. L. Barnes, *Transients in Linear Systems* (John Wiley and Sons, Inc., New York, 1942).

[13] An expression which is equivalent to $Z(R)$ has been given by J. C. Jaeger, Trans. Faraday Soc. **42**, 615 (1946).

[14] R. T. Bayard and D. Alpert, Rev. Sci. Instr. **21**, 571 (1950).

[15] D. Alpert, Rev. Sci. Instr. **22**, 536 (1951).

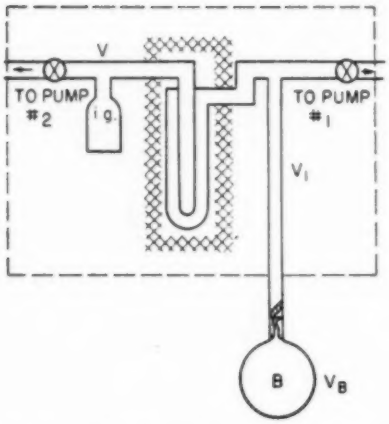


FIG. 4. Diagram of experimental vacuum system.

inner wall of the concentric cylinders shown enlarged and cut away in Fig. 3(b). The flow was radially inward. A schematic diagram of the system is shown in Fig. 4. The ion gauge shown on the right in Fig. 3(a) was used only to monitor the vacuum in V_1 during preparation of the vacuum system; it is omitted from Fig. 4. The gas to be studied was contained in the sealed supply flask B during the preparation of vacuum system. Reagent grade gas as supplied by Air Reduction Company was used without further purification. The part of the vacuum system inside the dashed line rectangle was enclosed in a large oven for overnight bakeout at 425°C before each experiment. During the actual experiments, the diffusion wall was enclosed in a furnace (below room temperature, a Dewar with an insulating cap) which is outlined by the cross-hatched area.

The furnace was made in two sections: base and cap. These joined snugly with overlapping flanges and with closely fitted openings to admit the vacuum leads and thermocouples. Base and cap units were separately controlled through Variacs from regulated ac lines. The temperature inside the furnace was maintained constant

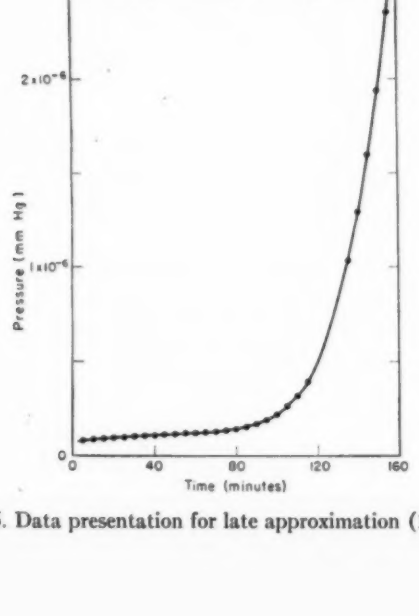


FIG. 5. Data presentation for late approximation (132°C).

and uniform within ±3°C as indicated by five thermocouples distributed over the glass inside the furnace. In the later experiments, the base unit was further stabilized by the addition of a special controlling unit which improved the temperature regulation to about ±1°C at temperatures below 250°C.

The ion gauge (i.g.) shown in Fig. 4 was operated from a self-regulating power supply unit which stabilized the electron emission current by controlling the filament temperature from a grid current error signal.[16] Ion collector current was measured with a bridge-type electronic microammeter and recorded automatically by a Leeds and Northrup Speedomax. Reading of the ion gauge was programmed by an automatic timer which operated the gauge and recorder chart for periods of fifteen seconds or less at regular intervals, which were preselected in multiples of one minute.

The final stage of vacuum system preparation consisted of outgassing the ion gauge electrodes by electron bombardment, while the regulated furnace was being brought up to the operating temperature. The high-vacuum valves were then closed and the final vacuum was attained by the pumping action of the gauge.[5] After the temperature had been established the constant rise rate due to leakage from the atmosphere was measured. The seal of the supply flask was then broken. The experimental data required for the determination of the diffusion coefficient, solubility, and permeability consisted of a record of the pressure indicated by the ion gauge as a function of the time elapsed after the breaking of the seal.

IV. CORRECTIONS TO EXPERIMENTAL DATA

Because the diffusion cell proper is inside the furnace during an experiment while the rest of the vacuum system is outside, a small correction for nonuniformity of temperature is required in interpreting the pressures in the system. In the volume V_1, where the pressure is always at least of the order of one mm Hg, the pressure is uniform throughout the volume, and would be correctly indicated by a manometer placed in that part of the volume which remains at room temperature. In these experiments, the pressure p_1 was calculated from the known pressure p_B which existed in the supply flask before the seal was broken. The resulting pressure in V_1 after expansion from the volume V_B of the flask is given by the formula

$$p_1=\frac{V_B}{V_1}\frac{p_B}{1-\alpha(1-T_r/T_f)}, \tag{19}$$

in which α represents the fractional part of V_1 which is at the temperature T_f inside the furnace. T_r represents the room temperature.

In the volume V on the low pressure side of the diffusion membrane, the pressure is always so low

[16] P. R. Malmberg, "Ionization Gauge Power Supply," Westinghouse Research Memo 411-9-16.

(below 10^{-3} mm Hg) that the mean free path in the gas is long compared to the diameter of the tubing which connects the ion gauge with the diffusion cell. In this case[17] the pressure p_r in the ion gauge is less than that inside the furnace in the ratio $(T_r/T_f)^{\frac{1}{2}}$. The readings of the ion gauge are corrected for this effect and also reduced to standard temperature by the formula

$$p=(273/T_r)p_r\{1-\beta[1-(T_r/T_f)^{\frac{1}{2}}]\}. \qquad (20)$$

in which β is the fractional part of V inside the furnace. In these experiments, α was approximately 0.1, the exact value depending on the size of the supply flask. The value of β was 0.261.

The determination of the diffusion coefficient in any of the cases discussed above clearly is not influenced in any way by the nonuniformity of temperature in the vacuum system since it does not depend on the numerical values of either p or p_1. However, the solubility and permeability calculations involve the ratio of numerical values of p to that of p_1, and therefore involve the product of the α-dependent and β-dependent temperature factors appearing in Eqs. (19) and (20). In computing the results which follow, these factors have been included as corrections after having obtained the k and S values from the plotted graphs of the uncorrected pressure data.

TABLE I. Experimental values of diffusion coefficient for helium in Pyrex (7740) glass at various temperatures and pressures evaluated by the early approximation method.

Temperature (°C)	Pressure, p_1 (mm Hg)	Diffusion coefficient (cm²/sec)
27	527	7.74×10^{-9}
204	543	4.52×10^{-7}
306	12.3	1.58×10^{-6}
132	11.8	1.41×10^{-7}
230	12.2	6.99×10^{-7}
87	537	4.63×10^{-8}

V. EXPERIMENTAL RESULTS

A graphical representation of typical data for helium taken from the recorder is shown in Fig. 5. This curve, if extended to later time, would approach the straight line of the late approximation [Eq. (17)]. Actually, in this case it was more convenient to use the early approximation [Eq. (14)]. By taking the slopes from Fig. 5, data were obtained for the plotting of Fig. 6. From Fig. 6 a value of D and a value of S were obtained by the early approximation method for the temperature specified. The values of D resulting from a series of measurements at different temperatures and pressures which were analyzed in this way are shown in chronological order in Table I.

The values of D obtained from these experiments were plotted on a logarithmic scale against the reciprocals of the absolute temperatures as shown in Fig. 7. The six values shown in Table I fit with a mean square deviation of 1.1 percent on a straight line which represents the relation

$$D=D_0e^{-T_0/T}. \qquad (21)$$

From the slope of the fitted line T_0 was found to be 3310°C, corresponding to an activation energy of 6.5×10^3 cal/mole, or 0.28 electron volt. D_0 was found to be 4.8×10^{-4} cm²/sec. The single value of D obtained from the measurement of diffusion through a thick-walled capillary lay about 10 percent below this line (4.64×10^{-7} cm²/sec at 215°C and 6.0 mm Hg).

The values of S obtained from these seven experiments using the early approximation [Eq. (16)] have an average of 0.0065 with a standard deviation of 58 percent. There is no clear indication of any dependence of S on temperature or pressure. A possible explanation

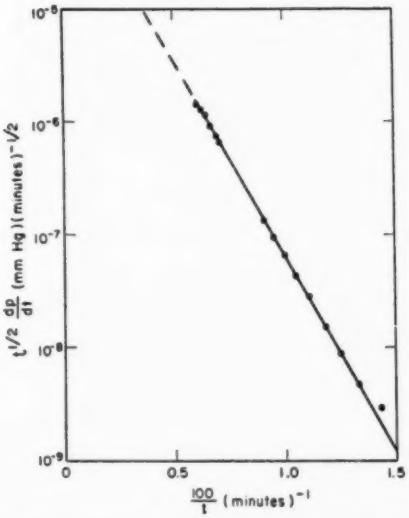


FIG. 6. Data presentation for early approximation (132°C).

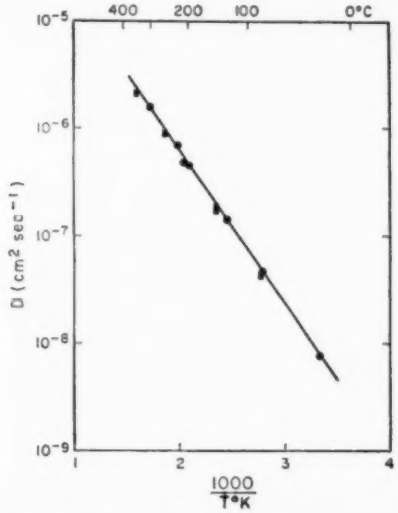


FIG. 7. Diffusion coefficient *vs* reciprocal temperature
o by early approximation method,
□ by late approximation method.

[17] E. H. Kennard, *Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1938), pp. 66, 330.

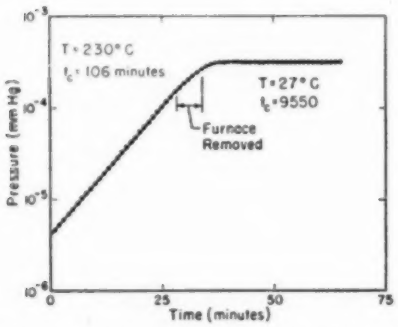


FIG. 8. Effect of a rapid transition from an established constant temperature to a new constant temperature. The new steady permeation rate is established immediately.

of the large deviations in S appeared to be in the nature of the calculation required by Eq. (16). The presence of D in the exponential in this equation makes the determination of S very sensitive to small errors in evaluation of D. In the hope of eliminating this particular difficulty, additional experiments were made for the specific purpose of determining S from the pressure intercept of the late approximation [Eq. (17)] which does not depend on the precision with which the numerical value of D has been calculated. The values of S determined by the latter method average 0.0079 with a standard deviation of 48 percent. The values of D resulting from these experiments are shown in Table II and Fig. 7.

Following the initial transient which gave the data for the determination of D and S, a steady state of flow was established. The permeability k was determined from the steady states which ensued after the six experiments referred to in Table I. In the measurements of permeability with a single value of the source pressure, it was observed that the steady rate of flow which followed diffusion at any given temperature could be established in a shorter time by allowing the diffusion to take place at a higher temperature and then switching rapidly to the lower temperature after the steady state was reached. The time required for the flow rate to achieve its new steady value was short in comparison with the diffusion time at even the highest temperature involved in the particular temperature transition. This effect is clearly illustrated in Fig. 8. The first part of the time interval represented there shows a measurement of the steady-state permeation at 230°C. During the time indicated on the figure the furnace was shut off and completely removed so that the glass cooled quickly to room temperature. The change of the diffusion current to its new constant value in only five minutes cannot possibly be due to a significant change in solubility because such a change would require adjustment of the concentration in the glass by diffusion, and the diffusion time l_c at 230°C is 106 min. In less than an hour enough points were obtained on a new scale to determine that the diffusion current was steady at the value which had been measured previously at room temperature. This result, which supports the belief that the solubility does not change with temperature by more than a few percent over the range investigated, was put to use in obtaining values of k at temperatures other than those at which diffusion measurements were made.

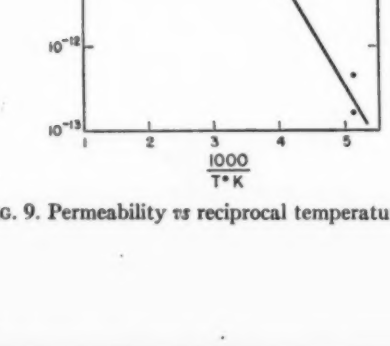


FIG. 9. Permeability vs reciprocal temperature.

TABLE II. Experimental values of diffusion coefficient for helium in Pyrex (7740) glass at various temperatures and pressures evaluated by the late approximation method.

Temperature (°C)	Pressure, p_1 (mm Hg)	Diffusion coefficient (cm²/sec)
353	0.73	2.1×10^{-6}
260	0.72	9.1×10^{-7}
150	1.0	1.9×10^{-7}
149	1.0	1.8×10^{-7}
83	1.0	4.4×10^{-8}

The results of the permeability measurements are presented in Fig. 9. The permeability, like the diffusion coefficient, is observed to vary with temperature according to the relation

$$k = k_0 e^{-T_0'/T}. \tag{22}$$

From the straight line fitted to the data in Fig. 9, the value of T_0' was found to be 3130°C, corresponding to an activation energy of 6.2×10^3 cal/mole or 0.26 electron volt. k_0 was evaluated as 2.0×10^{-6} cm²/sec. The equality of the activation energies for k and D within the experimental error provides additional evidence that the solubility, which is the ratio of k to D,[18] is not very sensitive to temperature.

Our values for permeability show good agreement throughout the whole temperature range with the values recently published by others[2] as is indicated in Fig. 10. The solubility as determined by the present experiments also agrees with the previously accepted value of 0.008.[3] To our knowledge there is no previously published directly measured value of the diffusion coefficient. Our values of activation energy for both the diffusion coefficient and permeability agree very well with those

[18] This relation between k, D, and S was shown in Eqs. (7) and (8).

obtained for the permeability by other recent experimenters. This suggests that the lack of agreement among the various permeability measurements may be due to differences in the solubility.

VI. CONCLUSIONS

An experimental method has been developed which is applicable to the study of the diffusion of a wide variety of gases through solids.[19] Using modern vacuum techniques and a straightforward analysis, the method permits simultaneous determinations of the diffusion coefficient, solubility, and permeability over a wide range of temperatures and pressures. The results of the experiments reported here for the case of helium in Pyrex indicate that the diffusion coefficient and permeability measurements are reproducible to within a few percent; although the direct measurements of solubility are less precise, the relation $DS=k$ is reasonably well verified and the solubility is shown to be quite insensitive to the glass temperature.

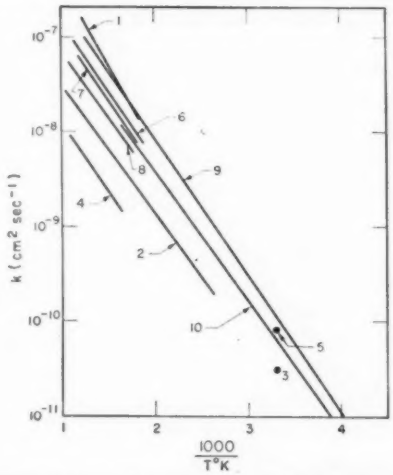


FIG. 10. Comparison of permeability values from various investigators (see footnotes 1 and 2): 1. Van Voorhis, 2. Urry, 3. Rayleigh, 4. Taylor and Rast, 5. Baxter, 6. Smith and Taylor (first sample), 7. Smith and Taylor (second sample), 8. Newkirk and Tooley, 9. Norton, 10. present experiments.

Experimental determination of the diffusion coefficient does not require any quantitative knowledge of gas pressures, but depends only on the linearity of the gauge used for the measurement of the gas diffusion current. Solubility and permeability can be determined from data which require only the calibration of this gauge relative to the manometer used for the measurement of the gas pressure in the source. Absolute calibration of pressure measuring instruments is therefore not necessary.

The analysis of data is such that sufficient data can be obtained in a short time for the determination of the diffusion coefficient at room temperature. The potential sensitivity of this method is indicated by the fact that it was possible in a preliminary measurement to make a fairly reliable measurement of D and S using only the atmospheric helium as a source.

[19] Two attempts to detect the passage of argon through Pyrex under a pressure of 540 mm Hg established the fact that at each of the two temperatures (390 and 450°C) the permeability for argon is less than 10^{-5} of that for helium at the same temperatures. By removing the atmospheric helium the limit of detection could be extended down several additional orders of magnitude.

APPENDIX

Units and Dimensions

Permeability values have appeared in the literature with several different systems of units, none of which are entirely satisfactory. Since the permeability is related to the diffusion coefficient and the solubility through the relation $k=DS$, it is convenient to clarify the dimensional nature of k by considering first the somewhat simpler and more generally accepted units of S.

The solubility has commonly been defined as "cc of gas/cc of glass/atmosphere." The quantity "cc of gas" appearing in the numerator of this expression refers to the volume which the dissolved gas would occupy in the free gaseous state at STP. An equivalent quantity is the pressure-volume product "cc-atmospheres" with the same standard temperature implied. On dividing this quantity by the volume of glass in cc one obtains the solute concentration in units of the concentration of free gas atoms at 0°C. The external gas pressure expressed in atmospheres is also an expression of concentration of free gas atoms at a specified temperature. Thus the solubility is the dimensionless ratio of concentration of solute inside the glass to that outside the glass.[20] The solubility values given in this paper conform to this interpretation. Stated without units, they may be understood as equivalent to units of "(pressure-volume of dissolved gas/volume of glass)/external gas pressure," where consistent units of pressure and volume may be chosen arbitrarily at the discretion of the reader.

Since the solubility is dimensionless, it follows that the dimensions and units of permeability are identical to those of the diffusion coefficient. The units of permeability (cm²/sec) given in this paper are equivalent to the expression (cm³ at STP/cm² sec)/(atmosphere of pressure difference/glass thickness of one centimeter) which is representative of the form used by earlier authors. If a permeability value in cm²/sec is inserted in Eq. (2) and the glass thickness d is expressed in cm, the current density j is given directly as the quantity of gas in cm³ at unit pressure flowing through each cm² in one second. The pressure unit is the same as that in which p_1 is expressed.

[20] We have followed accepted practice in expressing the quantity of gas dissolved in terms of gas pressure at the ice point, while expressing the external gas pressure (hence density) at the actual temperature of the furnace during the experiment. This has the effect of multiplying the solubility and also the permeability by the factor $273/T_f$. By converting the pressures in V to the equivalent pressures at the furnace temperature instead of the ice-point, one obtains effectively the correct numerical density of dissolved atoms in glass. The solubility would then mean the ratio of the actual particle density in glass to that in the gas outside the glass. The temperature would appear only as a parameter which describes the condition of the glass.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 7 JULY, 1954

Asymptotic Evaluation of the Field at a Caustic*

IRVIN KAY AND JOSEPH B. KELLER
Mathematics Research Group, New York University, New York, New York
(Received October 28, 1953)

The asymptotic expansion of wave fields in powers of $1/k$ for large k, where k is the wave number, has as its lowest-order term what is commonly known as the geometrical optics field. The caustics of geometrical optics are those point sets on which the zero-order terms become infinite. It is well known that caustics may exist even where the exact wave field is perfectly regular. An investigation of reflection from cylindrical walls of arbitrary cross section shows that the occurrence of caustic points means a change in character of the asymptotic expansion of the true field such that the lowest-order term is no longer independent of k but actually contains a factor k raised to a positive power. There also occurs a jump in phase along a ray passing through a caustic which, as is well known, equals $\pi/2$. As an application of the general method, the geometrical optics field is worked out in detail for the case of a plane wave incident on a parabolic cylinder, and the field is obtained in its lowest order at the focus and in the neighborhood of the focus. A similar problem for a reflector consisting of a segment of a circular cylinder is also considered in detail.

I. INTRODUCTION

THE electromagnetic field which can be constructed by the methods of geometrical optics, following R. K. Luneburg,[1] becomes infinite at caustics of the optical ray systems. In this paper we overcome this difficulty and obtain a finite value for the field at a caustic which is produced by the reflection of any incident wave from an arbitrary reflector (all in two dimensions for simplicity). What we actually determine is the leading term in the asymptotic expansion of the field with respect to $k=2\pi/\lambda$, for large k. Off the caustic this is just the geometrical optics field, but on the caustic it is proportional to $k^{1-1/n}$, where n is an integer greater than two. At a perfect focus, such as the focus of a parabolic cylinder, the field is proportional to $k^{\frac{1}{3}}$. Thus on a caustic or at a focus the field increases indefinitely as k does.

We have exemplified these general results by considering the special cases of a plane wave incident on the concave sides of a segment of a parabolic cylinder and a segment of a circular cylinder. In the parabolic case, with incidence along the axis of the parabola, the entire caustic is a point, the focus of the parabola. The asymptotic behavior of the field at the focus is obtained explicitly and transition formulae which show how the field varies in the neighborhood of the focus are also obtained. These formulas involve integrals which we have evaluated numerically in some special cases.

In the case of the circular cylinder, the caustic is a curve with a cusp. The asymptotic behavior at the cusp is different from that on the rest of the caustic. The variation of the field near the caustic is described in terms of Hardy's generalized Airy functions.

Since our results yield exactly the leading term in the asymptotic expansion of the field, they constitute a check on the solutions proposed by Debye,[2] Picht,[3] and Luneberg.[4] Each of these authors constructed a solution of the wave equation, or of Maxwell's equations, which is regular in an infinite homogeneous space and satisfies conditions at infinity. Luneburg proved that the solution thus obtained reduces to the known geometrical optics solution as k becomes infinite. In an actual problem involving boundaries the correct solution must satisfy certain boundary conditions. Since the solutions just mentioned do not satisfy such conditions, it is not certain that they will yield the correct field at a caustic in a real problem involving boundaries. However, we have proved that they do yield exactly the same asymptotic field on a caustic as is given by our method, and therefore their use in practical problems is justified.

II. FORMULATION: THE GEOMETRICAL OPTICS FIELD

We consider the solution $E(x,y)$ of the reduced wave equation

$$(\nabla^2+k^2)E=0. \tag{1}$$

The solution must satisfy the boundary condition

$$E=0 \quad \text{on} \quad C. \tag{2}$$

Here C is a given piecewise smooth curve which we will call a reflector.† At any corners of C the derivatives of E may become singular. Therefore we must impose a regularity condition at these corners, e.g., that the singularity be the weakest possible one, or that the "energy" integral be finite. In addition we assume that an incident field $E_i(x,y)$ which satisfies (1) is given, and that the reflected field $E_r=E-E_i$ satisfies the radiation

* This work was performed at Washington Square College of Arts and Science, New York University, and was supported in part by the U. S. Air Force, Air Force Cambridge Research Center, Air Research and Development Command.

[1] R. K. Luneburg, *Asymptotic Development of Steady State Electromagnetic Fields*, New York University, Washington Square College, Mathematics Research Group, Research Report No. EM-14 (1949).

[2] P. Debye, Ann. Physik **30**, 755 (1909).

[3] J. Picht, Ann. Physik **77**, 685 (1925).

[4] R. K. Luneburg, Asymptotic Evaluation of Diffraction Integrals, New York University, Washington Square College, Mathematics Research Group, Research Report No. EM-15 (1949).

† If C is closed, we consider the solution outside C only, and if C extends to infinity dividing the plane into two parts, we consider the solution in one part only.

condition

$$E_r = E - E_i \quad \text{is outgoing.} \tag{3}$$

These conditions complete the formulation of the problem. This formulation applies to the z component of the electric field when the reflector is a perfect conductor and the surrounding medium is homogeneous.

Rather than find E itself, we treat the simpler problem of determining the leading term in the asymptotic expansion of E with respect to k, for large k. This is what we call the geometrical optics field. We begin with the leading term in the asymptotic expansion of the incident field E_i. We assume it to be of the form

$$E_i \sim A_i(x,y) \exp[ik\psi_i(x,y)]. \tag{4}$$

Now, following Luneburg, we know that the zero-order term in the asymptotic expansion of the field is zero in the shadow. Furthermore, the leading term in the reflected field is

$$E_r \sim A_r(x,y) \exp[ik\psi_r(x,y)]. \tag{5}$$

The last statement must be qualified in two respects. First, the form (5) applies only at points not on caustics or shadow boundaries. Second, the leading term will be a sum of expressions of the type given in (5) at points through which more than one reflected ray passes. There will be one such summand for each ray.

If we apply (2) to the asymptotic form of E we obtain

$$E \sim A_i(x,y) \exp[ik\psi_i(x,y)] + A_r(x,y) \exp[ik\psi_r(x,y)] = 0 \quad \text{on} \quad C. \tag{6}$$

From (6) we have

$$A_r(x,y) = -A_i(x,y) \quad \text{on} \quad C \tag{7}$$

$$\psi_r(x,y) = \psi_i(x,y) \quad \text{on} \quad C. \tag{8}$$

(If E_r consists of a sum of terms, the remaining terms occur in pairs satisfying (7), (8), where the indices i and r refer to respective terms of the pair.)

A_r and ψ_r off C can be determined by the methods of goemetrical optics, using the values on C given by (7), (8). In particular, ψ_r satisfies the eiconal equation

$$|\nabla\psi_r| = 1. \tag{9}$$

Once ψ_r is found, A_r can be constructed by the conservation of energy formula

$$A_r(x,y) = \left(\frac{d\sigma_0}{d\sigma}\right)^{\frac{1}{2}} A_r(x_0,y_0). \tag{10}$$

Here x,y and x_0,y_0 are two points on a reflected ray, and $d\sigma_0, d\sigma$ respectively are cross-sectional widths of an infinitesimal tube of rays at each of these points. If we choose for x_0,y_0 a point on C, then (10) determines A_r off C. If D is the distance along the ray from x_0,y_0 to x,y and if K_r is the curvature of the reflected wave front at x_0,y_0, then it is easliy shown that

$$\left(\frac{d\sigma_0}{d\sigma}\right)^{\frac{1}{2}} = (1 - DK_r)^{-\frac{1}{2}}. \tag{11}$$

On the basis of Eqs. (5)-(11), the asymptotic form of the reflected field may be immediately written as

$$E_r(x,y) \sim -(1 - DK_r)^{-\frac{1}{2}} A_i(x_0,y_0) \times \exp[ik\psi_i(x_0,y_0) + ikD]. \tag{12}$$

The point x_0,y_0 is the point on C from which the ray through x,y is reflected, and D is the distance between these two points. If there is no reflected ray through (x,y), then $E_r(x,y) \sim 0$. When $DK_r = 1$, A_r becomes infinite. This occurs at the caustics of the reflected rays.

III. THE FIELD AT A CAUSTIC

In order to obtain a finite value for E_r at a caustic, we begin with Green's formula

$$E(x,y) = E_i(x,y) + \frac{1}{4i}\int_C \frac{\partial E}{\partial n} H_0^{(1)}(kD)\,ds. \tag{13}$$

Here D is the distance from the integration point x_0,y_0 on C to x,y, and $H_0^{(1)}$ is the zero-order Hankel function of the first kind. To obtain the asymptotic form of E at any point, and in particular at or near a caustic, we insert into the integrand the asymptotic form of $\partial E/\partial n$, which can be computed from (4) and (12). We also replace $H_0^{(1)}$ by its asymptotic form and then evaluate the integral asymptotically. In this way we get back the expression for E_r given by (12) at all points off the caustic, as we must expect. At the caustic, however, we obtain a finite result rather than the infinite value given by (12).

To justify the method described we must assume that the caustic does not intersect C. Furthermore the lower-order terms in the expansion of $\partial E/\partial n$, which have been neglected, must not contribute terms to the integral which are of the same (or higher) order as the contribution of the leading terms. That they do not can easily be proved to be the case if reasonable hypotheses are made about the omitted terms, but this will not be considered here.

We will now carry out the calculations just described. When $\partial E/\partial n$ is computed from (4) and (12) and the asymptotic form of $H_0^{(1)}$ is employed, Eq. (13) yields

$$E_r(x,y) \sim k^{\frac{1}{2}}(8\pi)^{-\frac{1}{2}} \int_C A_i \left[\frac{\partial\psi_i}{\partial n} - \frac{\partial\psi_r}{\partial n}\right] D^{-\frac{1}{2}} \times \exp\left[ik(\psi_i + D) - \frac{i\pi}{4}\right] ds. \tag{14}$$

This integral can be evaluated asymptotically by the method of stationary phase. The phase $\omega_i + D$ is a function of the integration point on C, and can be considered

as a function of arc length s along C. The points of stationary phase are determined by

$$\frac{d\psi_i}{ds}+\frac{dD}{ds}=0. \tag{15}$$

Then if $d^2/ds^2(\psi_i+D)\neq 0$ at the stationary point, (14) yields

$$E_r(x,y)\sim\frac{A_i}{2}\left(\frac{\partial\psi_i}{\partial n}-\frac{\partial\psi_r}{\partial n}\right)D^{-\frac{1}{2}}\left|\frac{d^2}{ds^2}(\psi_i+D)\right|^{-\frac{1}{2}}$$
$$\cdot\exp\left[ik(\psi_i+D)-\frac{i\pi}{4}\left\{1-\operatorname{sgn}\frac{d^2}{ds^2}(\psi_i+D)\right\}\right]. \tag{16}$$

In (16) all quantities on the right side are to be evaluated at the point of stationary phase determined by (15). If there is no stationary point the right side is to be replaced by zero; if there is more than one stationary point, the right side is to be a sum with one term of the above type for each such point.

It follows from (15) that at the stationary points either the law of reflection is satisfied or the reflected ray is the continuation of the incident ray. In the latter case, evaluation of (16) shows that $E_r\sim -E_i$ so that $E\sim 0$. This occurs in the shadow. In the former case (16) is found to coincide with (12), as is to be expected. It is to be noted that the phase jumps by $\pi/2$ across the caustic.

The expression (16) is not valid when $(d^2/ds^2)\times(\psi_i+D)=0$. But if for a point x,y this second derivative is zero and also the first derivative (15) is zero, then it can be shown as follows that the point lies on a caustic of the reflected ray. ψ_i+D is a function of x,y,s, say $g(x,y,s)$.

For each value of s the equation $g_s(x,y,s)=0$, which is (15), defines a ray. If $g_{ss}(x,y,s)=0$ also, then the point x,y lies on the envelope of the family of rays, which is the caustic.

In order to determine the location of the caustic, we compute $d^2/ds^2(\psi_i+D)$ and obtain

$$d^2/ds^2(\psi_i+D)=(D^{-1}+K_i)\cos^2\alpha-2K_e\cos\alpha. \tag{17}$$

Here K_i and K_e are the curvatures of the incident wave front and of the curve C, respectively, at the stationary point. The angle α is that between $\nabla\psi_i$ and the normal to C at the stationary point. The expression in (17) vanishes when D is given by (see also Keller and Keller[5])

$$D^*=(2K_e-K_i\cos\alpha)^{-1}\cos\alpha. \tag{18}$$

Thus (18) determines the distance D^* from C along a ray to the caustic, and thus the curvature K_r of the reflected wavefront at C is given by $K_r=(D^*)^{-1}$. The results (17), (18) are used in showing that (16) coincides with (12).

To obtain the field on the caustic we must evaluate the integral in (14) differently. For this purpose we consider the integral W given by

$$W=\int A(s)\exp[ikf(s)]ds. \tag{19}$$

Suppose that at a stationary point $\bar{s}$ at which $f'(\bar{s})=0$, we also have

$$f'(\bar{s})=f''(\bar{s})=\cdots=f^{(n-1)}(\bar{s})=0. \tag{20}$$

Then the Taylor series of $f(s)$ becomes

$$f(s)=f(\bar{s})+(s-\bar{s})^n\frac{f^{(n)}(\bar{s})}{n!}+\cdots \tag{21}$$

Therefore W is asymptotically given by

$$W\sim A(\bar{s})\exp[ikf(\bar{s})]\int\exp\left[ik\frac{f^{(n)}(\bar{s})}{n!}(s-\bar{s})^n\right]ds. \tag{22}$$

Thus we finally obtain

$$W\sim A(\bar{s})\exp[ikf(\bar{s})]2k^{-1/n}\Gamma(1+n^{-1})$$
$$\times(n!)^{1/n}\left|f^{(n)}(\bar{s})\right|^{-1/n}F_n, \tag{23}$$

where

$$F_n=\exp\left[\left(\frac{\pi}{2}i-n\right)\operatorname{sgn}f^{(n)}(\bar{s})\right]\quad\text{if } n \text{ is even, and}$$
$$F_n=\cos\frac{\pi}{2}n\qquad\text{if } n \text{ is odd.} \tag{24}$$

Let us apply the result (23) to the integral (14) when the point x,y lies on a caustic and when the first $n-1$ derivatives of ψ_i+D vanish at the stationary point. Then we have

$$E_r(x,y)\sim k^{\frac{1}{2}-1/n}(n!)^{1/n}\Gamma(1+n^{-1})(\pi D)^{-\frac{1}{2}}\sqrt{2}$$
$$\times\left|\frac{d^n}{ds^n}(\psi_i+D)\right|^{-1/n}\cos\alpha$$
$$\cdot\exp[ik(\psi_i+D)-i\pi/4]F_n, \tag{25}$$

where

$$F_n=\exp\left[\left(\frac{\pi}{2}i-n\right)\operatorname{sgn}\frac{d^n}{ds^n}(\psi_i+D)\right]\quad\text{if } n \text{ is even, and}$$
$$F_n=\cos\pi/2\qquad\text{if } n \text{ is odd.} \tag{26}$$

It is of interest to notice that the power of k which appears, namely $k^{\frac{1}{2}-1/n}$, is always a positive power on a caustic since $n\geq 3$. Thus the field increases indefinitely as k does. The larger the value of n the higher is the power of k, as one should expect because as n increases the phase of the integrand becomes more constant. If the

[5] J. B. Keller and H. B. Keller, Determination of Reflected and Transmitted Fields by Geometrical Optics New York University, EM-13, (1949) or J. Opt. Soc. Am. 40, 48–52 (1949).

phase is exactly constant, then the power $k^{\frac{1}{3}}$ occurs, as will be seen in the example of the parabolic cylinder.

IV. COMPARISON WITH THE DEBYE—PICHT—LUNEBURG SOLUTION

In their treatments of the field at a focus or caustic, Debye, Picht, and Luneburg each construct an exact solution of the wave equation, or of Maxwell's equations which is regular in the infinite homogeneous space. Debye, Picht, and Luneburg determine their solution by requiring it to have a specified behavior at infinity, and Luneburg proves that it must reduce to a specified geometrical optics field as k becomes infinite.

In most physical problems the field must satisfy boundary conditions on certain surfaces, and satisfy the wave equation or Maxwell's equations in a partly inhomogeneous medium. These conditions are not satisfied by the Debye-Picht-Luneburg solution. However, this solution does have the same behavior at infinity and the same geometrical optics field as the corresponding physical problem.

The question therefore arises as to whether the Debye-Picht-Luneburg solution implies that the field has the same asymptotic behavior at a caustic as that of the corresponding solution of the physical problem.

By the considerations of Sec. III (first and second paragraphs), both of these solutions, and in fact any solution with the correct geometrical optics field, will imply the correct behavior at caustics and foci provided that the lower-order terms in the integrand of (13) do not contribute higher-order terms. Thus all Debye-Picht-Luneburg solutions, although represented by less convenient integrals, will lead to the same leading asymptotic term at a caustic as does Eq. (14).

V. APPLICATIONS

1. The Parabolic Cylinder

Consider a cylinder, the cross section of which is a segment of a parabola (see Fig. 1). Let its focus be the origin of a Cartesian coordinate system, its axis be the x axis, and its equation be

$$y^2=-4p(x-p) \quad x\geq\alpha. \tag{27}$$

As is indicated in the above equation, the segment of the parabola under consideration lies to the right of $x=\alpha$ and is concave toward the left, since we assume that the focal length p is positive.

We now suppose that a plane wave is incident from the left with its direction of propagation parallel to the x-axis. We assume that this wave is given by

$$E_i=e^{ikx}. \tag{28}$$

The incident rays, which are the orthogonal trajectories of the incident wave fronts $x=$constant, are obviously the lines $y=$constant parallel to the x axis. Of these rays all those which interesect the parabolic segment will be reflected through the focus, and therefore the reflected wave fronts are segments of circles with the focus as origin. On the basis of this fact and Eq. (12), we have for the singly reflected field $E_r^{(1)}$

$$E_r^{(1)}(x,y)\sim-\left(\frac{D+r}{r}\right)^{\frac{1}{2}}e^{ik(2p-r)} \quad \text{before focus, and}$$
$$E_r^{(1)}(x,y)\sim-\left(\frac{D-r}{r}\right)^{\frac{1}{2}}e^{ik(2p+r)-i\pi/2} \quad \text{after focus.} \tag{29}$$

In Eq. (29), r is the distance of the point x,y from the focus and D is its distance along the reflected ray from the point of reflection on the parabola. The phase is easily determined since all rays reach the focus with the same phase $ik2p$.

Some rays intersect the parabolic segment a second time after having been reflected through the focus. In order to distinguish these rays, we introduce the angle θ between the once-reflected ray and the x axis. We also introduce the angle γ between the x axis and the line from the focus to the upper end of the parabolic segment. If $\gamma<\pi/2$ then $\alpha>0$ and the entire segment lies to the right of the focus. In this case no ray is twice reflected. On the other hand, if we consider the complete parabola, then $\gamma=\pi$ and every ray is twice reflected, except the ray incident along the x axis and characterized by $\theta=0$. In all other cases ($\pi/2<\gamma<\pi$) all rays are twice reflected except those satisfying

$$\pi-\gamma>\theta>-(\pi-\gamma), \tag{30}$$

which are once reflected.

After the second reflection, the reflected ray becomes parallel to the x axis ,and the doubly reflected field $E_r^{(2)}$ is found from Eqs. (12) and (29) to be given by

$$E_r^{(2)}(x,y)\sim 2p/ye^{ik(4p-x)-i\pi/2}. \tag{31}$$

The complete geometrical optics field can now be described as zero in the shadow, E_i at points outside the shadow through which no reflected ray passes, $E_i+E_r^{(1)}$ at points through which one reflected ray passes, $E_i+E_r^{(1)}+E_{r'}^{(1)}$ at points through which two singly reflected rays pass, $E_i+E_r^{(1)}+E_r^{(2)}$ at points through which one singly and one doubly reflected ray passes, and $E_i+E_r^{(1)}+E_{r'}^{(1)}+E_r^{(2)}$ at points through which two singly- and one doubly-reflected rays pass.

We now observe that the field $E_r^{(1)}$, as given by (29), becomes infinite at the focus of the parabola, and that the field $E_r^{(2)}$ as given by (31) becomes infinite along the axis $y=0$. This latter singularity appears only in the case of the complete parabola, since the doubly reflected ray $y=$constant suffers its first reflection at a point on the parabola with the ordinate $4p^2/y$. Thus as y approaches zero, the ordinate of the point of first reflection must become infinite, and this is possible only with the complete parabola. The x axis is not a caustic in the ordinary sense of an envelope of rays, but the geo-

metrical optics field becomes infinite there because the energy between incident rays with an arbitrarily large separation is reflected between rays with an arbitrarily small separation.

We will now assume that the parabolic segment is finite ($\gamma<\pi$), and we will obtain a finite expression for the field at and near the focus, which is the only caustic occurring in this case. To this end we employ (14), but we need not consider the contributions of $E_r^{(2)}$ to the integrand since the doubly reflected field contains no ray through the focus and therefore contributes only lower-order terms to the integral.

First we introduce the parabolic coordinates ξ,η defined by

$$\xi^2=(x^2+y^2)^{\frac{1}{2}}-x, \quad \eta^2=(x^2+y^2)^{\frac{1}{2}}+x \qquad x=\tfrac{1}{2}(\eta^2-\xi^2), \quad y=\xi\eta. \tag{32}$$

In these coordinates the equation of the parabolic segment is

$$\eta=\pm\sqrt{2p}, \quad 0\leq\xi\leq[2(p-\alpha)]^{\frac{1}{2}}. \tag{33}$$

Now (14) becomes

$$E_r^{(1)}(\xi,\eta)\sim-\left(\frac{kp}{\pi}\right)^{\frac{1}{2}}\exp\left(ikp-i\frac{\pi}{4}\right)\times\left\{\int_0^{[2(p-\alpha)]^{\frac{1}{2}}}[D^{-\frac{1}{2}}\exp(ikD-\tfrac{1}{2}\xi_0^2)]_{\eta_0=\sqrt{2p}}d\xi_0+\int_0^{[2(p-\alpha)]^{\frac{1}{2}}}[D^{-\frac{1}{2}}\exp(ikD-\tfrac{1}{2}\xi_0^2)]_{\eta_0=-\sqrt{2p}}d\xi_0\right\}. \tag{34}$$

At the focus this yields

$$E_r^{(1)}(0,0)\sim-2\sqrt{\frac{kp}{\pi}}\exp\left(2ikp-i\frac{\pi}{4}\right)\int_0^{[2(p-\alpha)]^{\frac{1}{2}}}\times(\tfrac{1}{2}\xi_0^2+p)^{-\frac{1}{2}}d\xi_0=-2\sqrt{\frac{2kp}{\pi}}\exp\left(2ikp-i\frac{\pi}{4}\right)\times\log\left[\left(1-\frac{\alpha}{p}\right)^{\frac{1}{2}}+\left(2-\frac{\alpha}{p}\right)^{\frac{1}{2}}\right]. \tag{35}$$

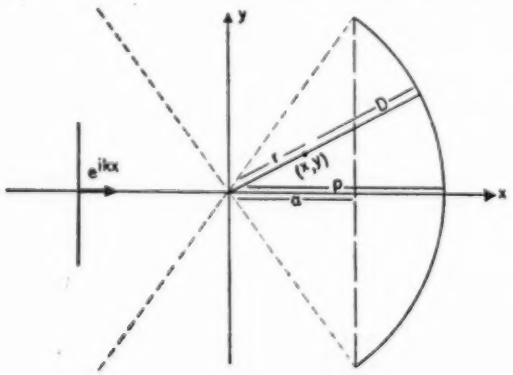


FIG. 1.

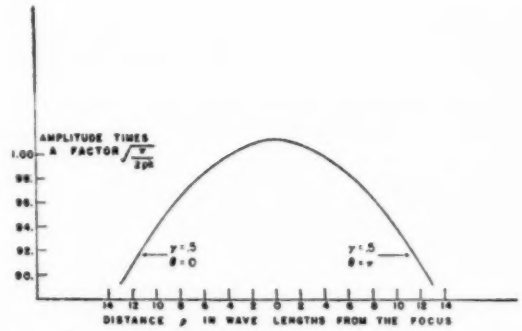


FIG. 2. Reflected field amplitude *versus* distance from the focus of the parabolic cylinder $\gamma=0.5$, $\theta=\pi$; $\gamma=0.5$, $\theta=0$.

To obtain the field near the focus we introduce the polar coordinates ρ,θ with the focus as origin and with ρ defined by $\rho=k(x^2+y^2)^{\frac{1}{2}}$. Thus ρ is 2π times the distance of a point from the focus measured in wavelengths. Now (34) becomes

$$E_r^{(1)}(\rho,\theta)\sim-\left(\frac{2kp}{\pi}\right)^{\frac{1}{2}}\exp\left(2ikp-\frac{i\pi}{4}\right)\times\int_{-\gamma}^{\gamma}\sec\frac{\phi}{2}\cdot e^{i\rho\cos(\theta+\phi)}d\phi. \tag{36}$$

Here $\gamma=\cos^{-1}[\alpha/(2p-\alpha)]$ is the angular coordinate of the end of the parabolic segment. Figures 2–4, based on Eq. (36), show the variation of the reflected field amplitude $|E_r^{(1)}|$ along and at right angles to the axis for two different parabolic segments.

2. The Circular Cylinder

A plane wave $\exp(ikx)$ is incident on the concave side of the sector of a circular cylindrical mirror of radius R. The situation is represented in Fig. 5. The center of the mirror has been placed at the origin of a Cartesian coordinate system so that the incident wave moves parallel to the x axis in a positive direction. The angular extent of the mirror is $\alpha+\beta$, where α measures the angle between the x axis and the end of the sector in the first quadrant, and β measures the angle between the x axis and the end of the segment in the fourth quadrant.

We shall consider only sectors such that no ray is reflected more than once. It is apparent from Fig. 5 that β can be at most $\pi-3\alpha$.

The phase of the incident wave at a point of reflection on the mirror is $R\cos\theta$, where θ is the angle of incidence; from Fig. 5 it is apparent that the equation of the reflected ray is

$$(R\sin\theta-y)/(R\cos\theta-x)=\tan2\theta; \tag{37}$$

(x,y) is an arbitrary point on the reflected ray.

Let D be the distance the reflected wave front has traveled from the point of reflection measured along the reflected ray. Then the phase ψ of the reflected wave front at the point of observation is $D+R\cos\theta$. It is clear

from Fig. 5 that

$$D=(R\cos\theta-x)/\cos 2\theta. \tag{38}$$

Thus the total phase at the point of observation is given by

$$\psi=R\cos\theta+(R\cos\theta-x)/\cos 2\theta. \tag{39}$$

By a similar consideration, or from (37), we obtain an expression for ψ involving y. Using this expression and (39) we obtain the following parametric equations for a reflected wave front i.e., a curve of constant phase ψ:

$$\begin{aligned} x&=(R\cos\theta-\psi)\cos 2\theta+R\cos\theta \\ y&=(R\cos\theta-\psi)\sin 2\theta+R\sin\theta. \end{aligned} \tag{40}$$

We can use (40) to obtain a coordinate system for points (x,y) on a reflected ray in terms of the distance D and the angle of incidence θ by replacing ψ by $R\cos\theta+D$:

$$\begin{aligned} x&=-D\cos 2\theta+R\cos\theta \\ y&=-D\sin 2\theta+R\sin\theta. \end{aligned} \tag{41}$$

Now we observe that the Jacobian of this transformation is $2D-R\cos\theta$, which is zero when $D=\frac{1}{2}R\cos\theta$. This is the distance D^* to the caustic given by our formula (18):

$$\begin{aligned} D^*&=(2k_c-k_i\cos\theta)^{-1}\cos\theta \\ &=(2/R)^{-1}\cos\theta=\tfrac{1}{2}R\cos\theta. \end{aligned} \tag{42}$$

We can check this result by calculating the curvature of the reflected wave front $\psi=$constant from (40); we obtain the result $K_r=2(R\cos\theta)^{-1}$ as expected. From our formula (12) we have for the reflected geometrical

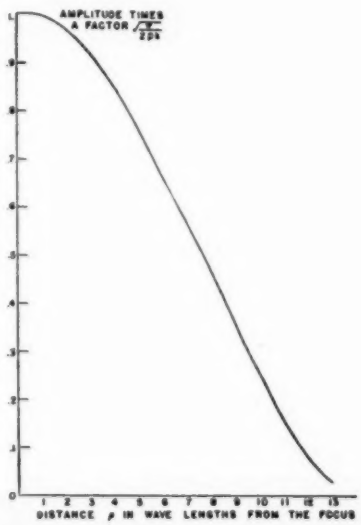


FIG. 3. Reflected field amplitude *versus* distance from the focus of the parabolic cylinder $\gamma=0.5$, $\theta=\pi/2$.

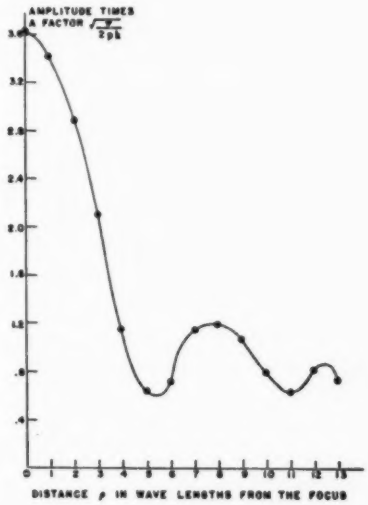


FIG. 4. Reflected field amplitude *versus* distance from the focus of the parabolic cylinder $\gamma=1.6$; $\theta=0$.

optics field the quantity

$$E_r\sim-\left(1-\frac{2D}{R\cos\theta}\right)^{-\frac{1}{2}}\exp[ik(R\cos\theta+D)] \tag{43}$$

in terms of the coordinate system defined by (41).

At the caustic, where $D=\frac{1}{2}R\cos\theta$, the reflected field takes a different form, which is found from (25):

$$\begin{aligned} E_r\sim&(48)^{\frac{1}{6}}\pi^{-\frac{1}{2}}\Gamma(4/3)\cos\frac{\pi}{6}\cdot k^{1/6}\sqrt{R\cos\theta} \\ &\times\left|\frac{d^3}{d\theta^3}(R\cos\theta+D)\right|^{-\frac{1}{6}}\exp\left[\tfrac{3}{2}ikR\cos\theta-i\frac{\pi}{4}\right]. \end{aligned} \tag{44}$$

To calculate the third derivative in this expression we make use of the fact that $d\psi/d\theta=0$ at the point of reflection and that $D=\frac{1}{2}R\cos\theta$ there. Then

$$r_0\sin(\theta-\theta_0)=\tfrac{1}{4}R\sin 2\theta. \tag{45}$$

In addition we use the fact that $d^2\psi/d\theta^2=0$ when the observation point is on the caustic, and obtain

$$r_0\cos(\theta-\theta_0)=R/2(1+\sin^2\theta). \tag{46}$$

Then, using (45) and (46), it is easy to find that when the point of observation is on the caustic,

$$\frac{d^3\psi}{d\theta^3}=-6R\sin\theta. \tag{47}$$

Thus the reflected field at the caustic is

$$\begin{aligned} E_r\sim&2\pi^{-\frac{1}{2}}\Gamma\left(\frac{4}{3}\right)\cos\frac{\pi}{6}\cdot(kR)^{1/6}\cos^{\frac{1}{6}}\theta \\ &\cdot\sin^{-\frac{1}{6}}\theta\exp\left[\tfrac{3}{2}ikR\cos\theta-i\frac{\pi}{4}\right]. \end{aligned} \tag{48}$$

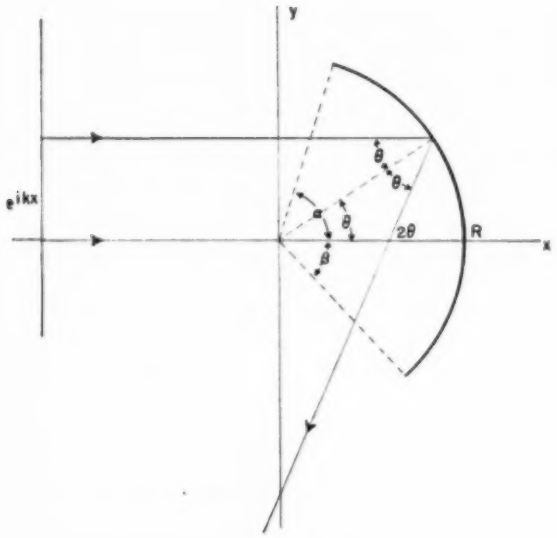


FIG. 5.

This result holds everywhere on the caustic except at $\theta=0$, where the caustic has a cusp. To obtain the field at this point we need one more derivative of ψ at the cusp; this is given by

$$\frac{d^4\psi}{d\theta^4}=-6R. \tag{49}$$

The reflected field at the cusp is then

$$E_r\sim 2^{\frac{1}{2}}\pi^{-\frac{1}{2}}\Gamma(5/4)\cdot(kR)^{\frac{1}{4}}\exp[3/2ikR-3/8i\pi]. \tag{50}$$

In the neighborhood of the caustic curve there is transition between the various forms of the reflected field given by (43), (48) and (50). To obtain this transition form we must return to the integral expression from which the general results were derived. In fact we write for the reflected field

$$E_r\sim(2\pi)^{-\frac{1}{2}}k^{\frac{1}{2}}D^{-\frac{1}{2}}R\cos\theta\exp[ik(R\cos\theta+D)=i\pi/4]\times\int\exp\Phi d\theta. \tag{51}$$

where

$$\Phi=\tfrac{1}{2}ik\{(D^{-1}R^2\cos^2\theta-2R\cos\theta)(\theta'-\theta)^2+\tfrac{1}{2}D^{-2}R^2\sin2\theta(D-R\cos\theta)(\theta'-\theta)^3\} \tag{52}$$

except in the neighborhood of the cusp. In the neighborhood of the cusp we have

$$\begin{aligned}\Phi=ik\{&[\tfrac{1}{2}D^{-1}R^2\cos^2\theta-R\cos\theta](\theta'-\theta)^2\\&+[\tfrac{1}{4}D^{-2}R^2\sin2\theta(D-R\cos\theta)](\theta'-\theta)^3\\&+1/24[R\cos\theta-D^{-1}R(R-D\cos\theta)\\&+4D^{-1}R^2\sin^2\theta-3D^{-3}R^3(R-D\cos\theta)^2\\&+18D^{-3}R^3\sin^2\theta(R-D\cos\theta)\\&-15D^{-3}R^4\sin^4\theta](\theta'-\theta)^4\}.\end{aligned} \tag{53}$$

If we use the value of Φ given by (53), the integral in (51) can be evaluated easily in terms of standard functions in the special case where the observation point is on the ray reflected through the cusp, the ray being given by $\theta=0$. On the other hand, if we use the value of Φ given by (52), we can evaluate the integral in (51) with no restriction.

Consider the integral in (51) with Φ given by (52). It has the form

$$I=\int_{-\infty}^{\infty}\exp[i(\gamma t^3+\delta t^2)]dt=|\gamma|^{-\frac{1}{3}}\int_{-\infty}^{\infty}\exp[i(s^3+\delta\gamma^{-\frac{1}{3}}s^2)]ds. \tag{54}$$

Let $s=u-\delta\gamma^{-\frac{1}{3}}/3$. Then

$$\begin{aligned}I&=2|\gamma|^{-\frac{1}{3}}\exp\left[\frac{2i\delta^3}{27\gamma^2}\right]\int_0^{\infty}\cos\left(\frac{u^3-\delta^2u}{3\gamma^{4/3}}\right)du,\\&=2|\gamma|^{-\frac{1}{3}}\exp\left[\frac{2i\delta^3}{27\gamma^2}\right]Ci_3\left(\frac{-\delta^2}{9\gamma^{4/3}}\right),\end{aligned} \tag{55}$$

where $Ci_n(a)$ is Hardy's generalized Airy integral given by Watson:[6]

$$Ci_3(a)=\left(\frac{\pi\alpha^{\frac{1}{2}}}{6\sin(\pi/6)}\right)[J_{-\frac{1}{3}}(2\alpha^{\frac{3}{2}})-J_{\frac{1}{3}}(2\alpha^{\frac{3}{2}})]. \tag{56}$$

According to (52) we have

$$\begin{aligned}\gamma&=(kR^2\sin2\theta/4D^2)(D-R\cos\theta)\\\delta&=\tfrac{1}{2}k[(R^2\cos^2\theta/D)-2R\cos\theta],\end{aligned} \tag{57}$$

and for the reflected wave field in the neighborhood of the caustic

$$E_r\sim(2k/\pi D)^{\frac{1}{2}}|\gamma|^{-\frac{1}{3}}R\cos\theta\exp\left[ik(R\cos\theta+D)+i\left(\frac{2\delta^3}{27\gamma^2}-\frac{\pi}{4}\right)\right]Ci_3\left(\frac{-\delta^2}{3\gamma^{4/3}}\right). \tag{58}$$

It is understood of course that (58) holds only when the point of observation is away from the cusp.

If the point of observation is in the neighborhood of the cusp and on the line $\theta=0$, we can use (51) and (53) to obtain the reflected wave field in a form similar to (58). Here the integral on the right of (51) has the form

$$I=\int_{-\infty}^{\infty}\exp[-i(\alpha\theta'^4+\beta\theta'^2)]d\theta'=\alpha^{-\frac{1}{4}}\int_{-\infty}^{\infty}\exp[-i(t^4+\beta\alpha^{-\frac{1}{2}}t^2)]dt. \tag{59}$$

[6] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (The Macmillan Company, New York, 1948), p. 320 ff.

We assume $\alpha>0$ in the region considered, since $D\sim\frac{1}{2}R$ there. Then

$$I=a\alpha^{-\frac{1}{4}}\exp(i\beta^2/8\alpha)\{C_{i4}(\tfrac{1}{4}\beta\alpha^{-\frac{1}{2}})-iS_{i4}(\tfrac{1}{4}\beta\alpha^{-\frac{1}{2}})\}. \quad (60)$$

Thus

$$E_r\sim(2k/\pi D)^{\frac{1}{2}}R\alpha^{-\frac{1}{4}}\exp[ik(R+D)+i(\beta^2/8\alpha-\pi/4)]\{C_{i4}(\tfrac{1}{4}\beta\alpha^{-\frac{1}{2}})-iS_{i4}(\tfrac{1}{4}\beta\alpha^{-\frac{1}{2}})\}, \quad (61)$$

with

$$\alpha=\frac{1}{24}[RD^{-1}(R-D)-R+3R^3D^{-3}(R-D)^2] \quad (62)$$

$$\beta=R-\tfrac{1}{2}R^2D^{-1}.$$

Again the functions C_{i4} and S_{i4} are Hardy's generalized Airy integrals:[6]

$$C_{in}(\alpha)=\pi\alpha^{\frac{1}{2}}[2n\sin(\pi/2n)]^{-1}\{J_{-1/n}(2\alpha^{n/2})-J_{1/n}(2\alpha^{n/2})\}$$
$$S_{in}(\alpha)=\pi\alpha^{\frac{1}{2}}[2n\cos(\pi/2n)]^{-1}\times\{J_{-1/n}(2\alpha^{n/2})+J_{1/n}(2\alpha^{n/2})\}. \quad (63)$$

Equation (61) gives the reflected field in the neighborhood of the cusp provided that the observation point is on the axis of the circular mirror.

ACKNOWLEDGMENTS

The authors wish to thank Professor Fritz Reiche for his stimulating discussions during this investigation.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 7 JULY, 1954

Secondary Emission from Nichrome V, CuBe, and AgMg Alloy Targets Due to Positive Ion Bombardment

MICHAEL J. HIGATSBERGER,* H. L. DEMOREST, AND ALFRED O. NIER
University of Minnesota, Minneapolis, Minnesota
(Received November 19, 1953)

A mass spectrometer ion collector system has been devised for the investigation of secondary electron emission by impact of ions on surfaces. Results are reported for various ions incident on clean, baked (but not atomically clean) surfaces of AgMg, CuBe, and Nichrome V targets. For rare gas isotopic ions it is found that for a given energy ion the value of γ varies approximately inversely with the square root of the mass.

INTRODUCTION

WITH the increased use of electron multipliers as sensitive detectors for positive ions in devices such as mass spectrometers, a detailed knowledge of the process of creation of secondary particles by ion bombardment is desirable. Previous investigations have followed closely the general experimental technique used by Oliphant.[1] Massey and Burhop[2] have recently published a summary of this field. Hagstrum[3] has just published the results of a comprehensive study of electron ejection from molybdenum as a result of helium bombardment. Because secondary emission is affected by many parameters, such as mass, energy, and type of incident ion as well as the condition and type of surface, it is profitable among other things to study isotopic ions in order to learn something about mass dependence.

APPARATUS

(a) Mass Spectrometer

The ions in the present investigation are produced through electron impact and separated according to mass by a single-focusing 60° mass spectrometer.[4] The resolving power $\Delta M/M$ was approximately 1/250. The ion accelerating voltage was variable between 500 and 6000 volts, and short term stability was one part in 5000. The spread in energy of the ion beam was estimated to be less than two volts. The electron beam was held constant by using a space charge emission regulator;[5] the ion beam was steady and free from drift. For detecting the ion current as well as the secondary emission current, a commercially available portable battery-operated vacuum tube electrometer[6] was used. The voltage drop across a high input resistor (10^9, 10^{10}, 10^{11} ohms) is applied to the grid of a subminiature electrometer tube. Because the instrument is compact and portable, it could be insulated from ground and operated at elevated potentials.

* Fulbright Fellow. Present address: Physics Institute, University of Vienna, Vienna, Austria.

[1] M. L. Oliphant, Proc. Roy. Soc. (London) **A127**, 373 (1930).

[2] H. S. W. Massey and E. H. S. Burhop, *Electronic and Ionic Impact Phenomena* (Oxford Press, London, 1952), p. 541.

[3] H. D. Hagstrum, Phys. Rev. **89**, 244 (1953).

[4] A. O. Nier, Rev. Sci. Instr. **18**, 398 (1947).

[5] E. B. Winn and A. O. Nier, Rev. Sci. Instr. **20**, 773 (1949).

[6] Model 200, Keithly Instrument Company, Cleveland, Ohio.

The high voltage supply for the deflecting plate of the collector assembly described in the next section is variable from 500 to 12 000 volts and is stable to 1 part in 5000.

(b) Collector System

Figure 1 shows schematically the collector system. Ions pass through the slit in P after being separated according to mass in the 60° magnetic sector field. The focus produced is in the plane P. Secondary electrons produced on the edges of the slit in P will be repelled by the plate R held at a potential of -90 volts, and hence do not reach the collector C_1. The target plate T and the deflecting plate D form a parallel plate condenser. T is always grounded and in a fixed position. The potential of D can be varied. When measuring the primary ion current, D is at ground potential and allows the beam to enter the V shaped collector box C_1. Since the collector is 45 volts positive with respect to D, secondary electrons produced in it will not be able to leave. The V shape of the collector boxes makes it unlikely that charged particles reflected from the walls of the boxes will leave. D, with the attached collector assembly and surrounding shield box, is movable parallel to T as the arrows indicate. The construction is such that the position can be adjusted from the outside through a sylphon bellows. It was not necessary to turn off the voltage while adjustment was made.

The application of a positive potential on D will deflect the ions in a parabolic path and if a large enough potential is chosen, the ions will finally fall back against target T producing secondary emission. The negatively charged secondary particles follow the field lines of the electric field between T and D and are collected in C_2. In this way separation between positive and negative particles is obtained. The energy of the ions striking the target is independent of the potential on D. The trajectory of the ion path changes with the potential on D. Therefore, it is possible to study the secondary emission along the target T. If the ion beam emerging through the slit in T makes an angle α with the normal to T and if "a" is the separation between T and D, the deflected ions will strike T at a distance x from the slit in T where $x=2(V_A/V_D)a \sin 2\alpha$. V_A and V_D are respectively the accelerating voltage given to the ions in the spectrometer ion source and the voltage applied to the deflecting plate D. Ions strike at an angle α with the normal, independent of the value of V_D. The maximum x coordinate is obtained when $V_A \cos^2\alpha = V_D$. This is the case when the ions just graze the plate D before returning to T. In our special case $a=1$ cm and $\alpha=45°$. The ion beam striking the target can be moved along T from $x=4$ cm for $V_A/V_D=2$ to $x=0$ for $V_A/V_D=0$.

The paths of the electrons which are emitted from the target are not sensitive to the weak stray magnetic field of the spectrometer. If a magnetic field is applied parallel to the deflecting plate and in the plane of Fig. 1, it is possible to measure the percentage of negative ions formed on the target with respect to the number of electrons. Under such field conditions the electrons would move perpendicular to the plane of the paper in a cycloid starting from the point where the primary ions impinge. It may be shown that if H is the magnetic field and $E=V_D/a$, then if the fields satisfy the equation $a \geq 2Em/eH^2$, the electrons would not be collected. On the other hand, the negative and positive ions, because of their much greater mass, would not be affected appreciably.

If we summarize the advantage of the described collector system, we find:

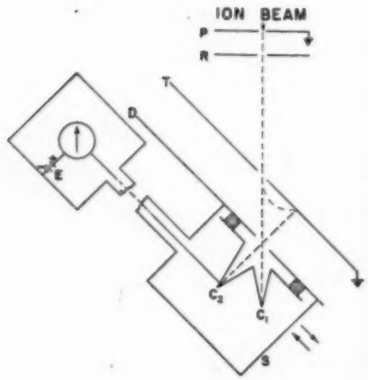


FIG. 1. Schematic drawing of collector system. Slits in P, R, T, and D have dimensions of 0.25×6, 1×10, 1×10, and 2×10 mm, respectively. Collectors C_1 and C_2 are supported from plate D by insulators as shown and are surrounded by shield box S. Continuation of shield S surrounds the electrometer amplifier and lead. This construction eliminates leakage currents to ground in the measuring circuit when collector arrangement is at high potential.

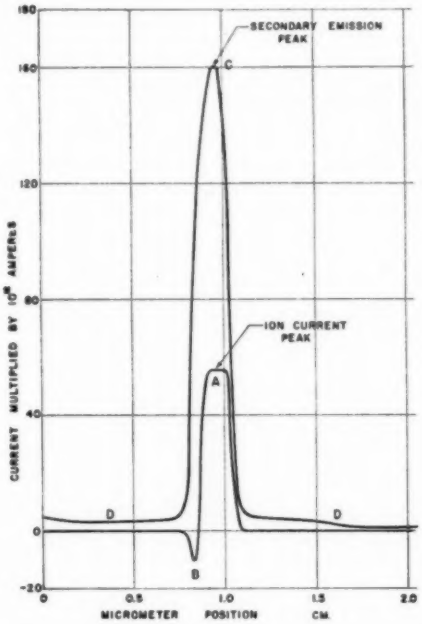


FIG. 2. Ion current and negative secondary emission current from target T as a function of position of collector assembly. Zero position of scale occurs when perpendicular dropped from slit in T strikes slit in D.

(1) Complete separation between primary and secondary particles, as has been shown in previous work.[7]

(2) Fair independence of stray magnetic fields.

(3) No additional secondary emission due to uncharged particles which may accompany the primary ion beam.

Targets of Nichrome V, CuBe (98 percent Cu, 2 percent Be), and AgMg (98.3 percent Ag, 1.7 percent Mg) were studied. They were prepared by polishing the surface on a brush polisher until no interference colors were visible. This was followed by a washing with benzine, acetone, alcohol, and distilled water. The dry target was then placed on the high vacuum system and heated for 100 hours to a temperature between 220 and 250°C as the pressure in the system dropped from an initial value of 5×10^{-6} to 10^{-7} mm Hg.

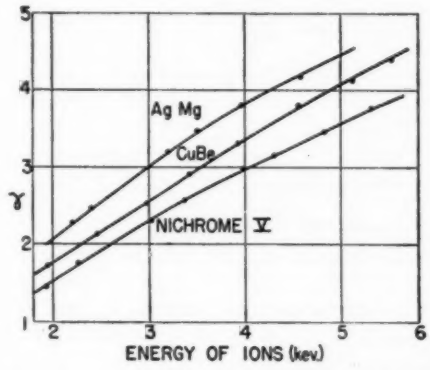


FIG. 3. Secondary emission factor γ for A^+ ions of various energy striking targets of AgMg, CuBe, and Nichrome V alloys. In the present paper γ is defined as the ratio of the number of secondary electrons to the number of incident ions.

The targets thus prepared showed constant reproducible secondary emission factors over a period of weeks so long as the vacuum was maintained. Opening the system to air and repumping in general resulted in a change in the value of the factors.

MEASUREMENTS

When measuring the ion current of an isotope, the plate D is at ground potential and in such a position that the ions strike the collector C_1. The micrometer adjustment moving the collector system allows a 5-cm movement of the plate D. In Fig. 2 we see the ion current collected in C_1 as slit in D is moved relative to slit in T. The ion current remains zero until we have moved plate D, 0.75 cm from our starting position. Here a negative current is observed, point B. The current then goes positive and returns to zero at 1.1 cm. The negative peak is caused by ions striking the right edge of slit D and producing secondary electrons which,

[7] Michael J. Higatsberger, Z. Naturforsch. **6a**, 151 (1951); **8a**, 206 (1953).

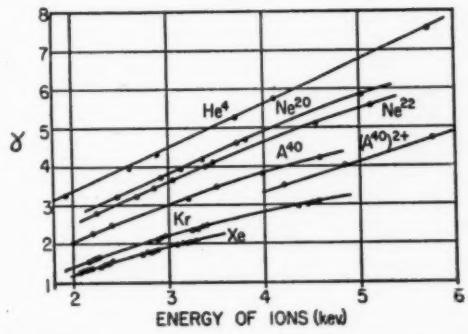


FIG. 4. Secondary emission factor γ for various rare gas ions striking an AgMg target.

under the influence of the potential, fall on the collector. A corresponding effect does not occur on the other side of the slit because of the beveling of the slit edges. We see that the peak A stays constant over a certain range. For assurance that we collect all the ions during a measurement, the measurements are made in the middle of peak A.

Upon applying a positive potential on the deflecting plate D, the primary ions are bent and hit the target T, producing secondary emission. Holding the deflecting voltage constant and moving the split D from position 0 to 2 cm, we have the upper curve CD in Fig. 2. The peak C appears constant over the position where the ions strike the target. When the ions are bent around and strike the target, some of them will be reflected from the target surface. These ions, behaving as if they originated at the point of reflection, will, in turn, be deflected by the electric field. They will strike the target at a point determined by the angle and energy they have after reflection. The tail D on the secondary emission curve is produced by these ions. Presumably a careful analysis of the shape of the tail would give information concerning the angles and energies of the reflected ions.

When the ion beam was moved along a baked Nichrome target for 3 cm, the secondary emission factor remained constant. CuBe and AgMg gave slight variations which were assumed to be due to inhomogeneities in the target surface. If this assumption is true, we have a sensitive test for surface homogeneity of metals.

Measurements were made employing singly charged ions of the principal isotopes of the noble gases. Also, investigations of secondary emission were made using doubly charged particles and some simple molecular ions and fragment ions. Figures 3, 4, and 5 show some of the results obtained. The measurements were taken at the peak position, and no attempt was made to measure the percent of reflected ions. The points on the curves are accurate to about 1 percent the accuracy of the detecting instrument. The experimental points shown are individual measurements and not the average of many runs. Over a period of several weeks, during which time the vacuum was maintained, spot checks indicated the results were reproducible within 1 percent.

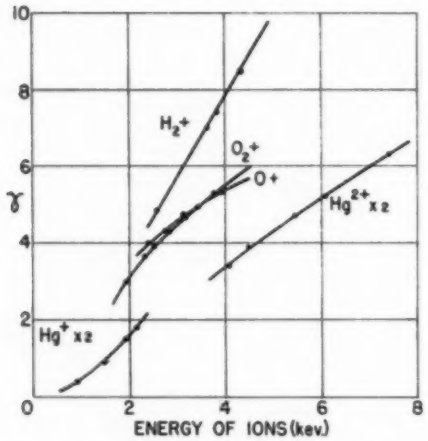


FIG. 5. Dependence of secondary emission factor γ upon ion energy for H_2^+, O^+, O_2^+, $(Hg^{202})^+$ and $(Hg^{202})^{2+}$ striking an AgMg target. Data for the other singly charged Hg isotopes were taken but in interest of simplicity are not shown in figure. However, within experimental error of measurements, the same empirical isotope mass correction as found for rare gas ions causes points for the several isotopes to lie closer to the smooth curve drawn than if no correction is made.

DISCUSSION

In the present investigation V_D/V_A was held at a value of 2 at all times. Hence, ions always struck the same region of the target. Figure 3 shows the secondary emission for A^{40} ions striking on baked Nichrome V, CuBe, and AgMg alloys. Curves similar in appearance are obtained for other ions striking targets of these three alloys. A comparison of the data for AgMg and CuBe shows that at a given energy the ratio of the secondary emission factors for AgMg and CuBe is very nearly constant for all of the impinging ions employed. When the results for Nichrome V are compared, close agreement is found in all cases except for the helium ions. Because the value of γ depends upon the treatment of the surface, no significance should be attached to the relative position of the curves in Fig. 3. In the case of targets which were not baked but only polished and washed in the manner earlier described, the secondary emission factor was considerably smaller and varied with time.

Figure 4 shows the variation of the secondary emission factor with the accelerating potential applied to the ions for the principal isotopes of He, Ne, A, Kr, and Xe striking a baked AgMg target. For the isotopes of neon we find well separated emission curves. Ploch[8] and Ploch and Walcher[9] investigated the secondary emission produced by isotopic ion pairs Li^6-Li^7, $Ne^{20}-Ne^{22}$, and $K^{39}-K^{41}$, and concluded that in the energy range 600–6000 volts for isotopic ions equal velocities gave within the limit of error of 1 percent equal γ values. We have not been able to confirm this observation. However, we have shown our Ne^{20} and Ne^{22} curves can be brought into coincidence if it is assumed that for a given energy of ion the γ value varies inversely with the square root of the mass.† Because the percentage variation in the masses of the krypton and xenon isotopes is not as great as for the neon isotopes the mass effect is not so apparent for these elements. However, if the same correction is assumed here, the points for the several isotopes will, in each element, lie along a single smooth curve. The points in Fig. 4 are not corrected for the isotope mass effect. However, on a carefully drawn large drawing, the effect can be seen very clearly. In fact, except for helium, the rare gas curves themselves agree approximately if they are related by the square root of the mass ratio factor. The data for the mercury isotopes (Fig. 5) is also improved by this correction.

Figure 4 also shows the dependence of the secondary emission factor upon ion energy for argon atoms which have had two electrons removed prior to acceleration. The curve lies slightly below the corresponding one for singly charged ions. Inghram, Hayden, and Hess[10] report the curves for doubly charged ions to lie slightly above those for singly charged.

Figure 5 gives the comparison for several miscellaneous ions. From the data here as well as in Fig. 4, it is clear that for a given energy, lighter ions in general give a relatively larger γ. As pointed out by Inghram *et al.*,[10] this is modified by the electronic structure of the ion.

In Hagstrum's[3] recent work he finds that for low-energy (below 1000 volts) He^+, He^{2+}, and He_2^+ ions incident upon an atomically clean Mo surface, the value of γ is almost independent of ion energy and always less than 1. Offhand, the present results are difficult to reconcile with his findings. In a private communication to one of us (A.O.N.) he reports that for noble gas ions incident on tungsten surfaces which are not atomically clean, γ approaches zero as in the present work.

The present work was undertaken to gain empirical data on the applicability of secondary electron multipliers as ion detectors in mass spectroscopy. However, the method of investigation appears to lend itself to systematic studies of the collision of ions with surfaces. The large range of ion energy possible will enable one to study regions of both potential and kinetic ejection. The high sensitivity and flexibility of the mass spectrometer makes possible studies on the dependence of γ on ion mass and chemical form. With a modification of the apparatus, a simultaneous investigation of the energy of the ejected electrons will be possible.

This research was made possible by a grant from the American Cancer Society acting through the Committee on Growth of the National Research Council.

[8] W. Ploch, Z. Physik **130**, 174 (1951).

[9] W. Ploch and W. Walcher, Rev. Sci. Instr. **22**, 1028 (1951); "Mass Spectroscopy in Physics Research," NBS Circular 522 (U. S. Government Printing Office, 1953), p. 255.

† In a new investigation using the same apparatus Mr. J. W. Greenwood in this laboratory has confirmed this observation and verified it also for the isotopic pair $A^{36}-A^{40}$.

[10] Inghram, Hayden, and Hess, "Mass Spectroscopy in Physics Research," NBS Circular 522 (U. S. Government Printing Office, 1953), p. 257.

Information Theory and Uncertainty Principle

L. BRILLOUIN
International Business Machines-Watson Laboratory, New York
(Received November 18, 1953)

The methods of information theory are used in the discussion of observations under a microscope. Conclusions are similar to those obtained in a previous paper and exemplify again the difficulty of measuring extremely small distances. A comparison with the discussion of uncertainty relations shows that quantum uncertainties can be very much higher than the theoretical limit, especially when material screens or shutters are not being used and the observation is made on a free particle. This correction to uncertainty relations, however, does not introduce any change in the theory of quantized fields of Bohr and Rosenfeld.

I. INTRODUCTION

THE negentropy principle of information was stated and discussed in a previous paper.[1] The statistical definition of "information" was given as

$$I=k\ln\frac{P_0}{P_1}; \tag{1}$$

P_0, initial number of equally probable possibilities, before any information, P_1, final number of possibilities, after the information I has been obtained, ln, natural logarithm, $k=1.38\,10^{-16}$, Boltzmann constant, thermodynamical units CGS centigrade degrees.

It was shown that information corresponds to a negative term in the total entropy of a system. A number of special cases of application were discussed. They exemplified how an observation, giving an information ΔI must always be paid for by a larger increase ΔS in entropy. The efficiency ϵ of the experimental method was defined by

$$\epsilon=\frac{\Delta I}{\Delta S}\leqslant 1. \tag{2}$$

Some practical applications of these definitions were discussed.[2]

We now want to investigate in detail the problem of observations under a microscope. The general situation was briefly sketched in the preceding paper (see reference 1, sec. 8, p. 1160) and a careful investigation is of interest, because it throws a new light on the usual discussion of uncertainty relations:

$$\Delta x\Delta p_x\sim h. \tag{3}$$

We shall prove that in some cases the limit h cannot be reached and that the fundamental uncertainty is very much greater,

$$\Delta x\Delta p_x\sim Ah,\quad A\gg 1. \tag{4}$$

The numerical factor A depends upon experimental conditions, and may reach very large values up to high powers of 10.

The problem of quantized fields is discussed and it is shown that the new condition (4) does not modify the usual conditions and that the quantum uncertainties of Bohr-Rosenfeld remain valid.

[1] L. Brillouin, J. Appl. Phys. **24**, 1152 (1953).
[2] L. Brillouin, J. Appl. Phys. **25**, 595 (1954).

II. INFORMATION AND OBSERVATION UNDER A MICROSCOPE

In all problems previously discussed it was found necessary to specify the total field of observation, in order to obtain a definition of the information. We shall now consider the problem of observing the position of a particle over a plane rectangular area ab, with errors Δx and Δy in the xy directions. The total time interval for one observation is τ and the error in the definition of an instant of time is Δt. The whole initial uncertainty was $ab\tau$, but after the observation is made, it is reduced to $\Delta x\Delta y\Delta t$. These two quantities correspond to P_0 and P_1 in Eq. (1) and we define the information obtained

$$\Delta I=k\ln\frac{ab\tau}{\Delta x\Delta y\Delta t}. \tag{5}$$

Let us now consider experimental conditions of observation with a microscope. We may think of light coming up through a wave guide of cross section ab (Fig. 1). The wave guide is being used to define the total area of observation. A particle M scatters light on a lens L which gives an image of M at I. We want to discuss the number of photons required to give the required accuracy and to define the position of M within Δx, Δy, and Δt.

The wave guide obtains a discrete set of proper waves.[3] The problem is to find out how a superposition of such waves can result in a focus at a given point M, at a certain time t.

Let us select the H (transverse electric) waves with a vertical magnetic field whose magnitude must have a zero normal derivative on the boundary of the wave guide:

$$H_z=\sum A_{lm}\cos\frac{\pi lx}{a}\cos\frac{\pi my}{b}\cos\omega\left(t-\frac{z}{V_{lm}}\right). \tag{6}$$

l and m represent the number of nodes of a certain wave in the x and y direction, while V_{lm} is the phase velocity

[3] See, for instance, Massachusetts Institute of Technology Radiation Laboratory Series, **8**, 34, 37, 55. Montgomery, Dicke, and Purcell, *Principles of Microwave Circuits* (McGraw-Hill Book Company, Inc., New York, 1948).

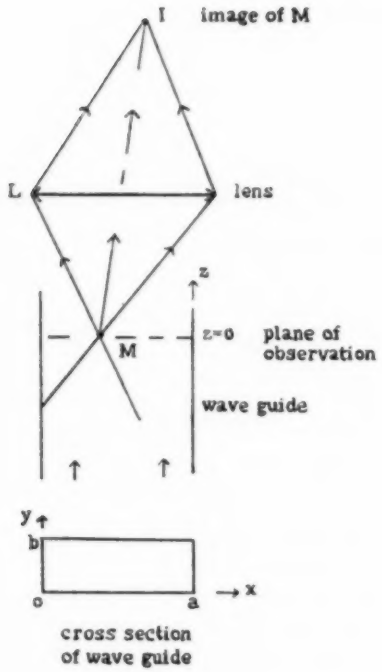


FIG. 1.

of this particular wave in the z direction. We limit our discussion to the case of H waves, but similar considerations apply for E waves (transverse magnetic). The integers l, m can run from 0 to ∞ but the combination 0, 0 is excluded. The wave propagation equation

$$\nabla^2 H_z - \frac{1}{c^2}\frac{\partial^2 H_z}{\partial t^2} = 0, \tag{7}$$

yields

$$\omega^2\left(\frac{1}{c^2} - \frac{1}{V_{lm}^2}\right) = \left(\frac{\pi l}{a}\right)^2 + \left(\frac{\pi m}{b}\right)^2. \tag{8}$$

The phase velocity V_{lm} is larger than c, and the group velocity $U_{l,m}$ is smaller than c. Both velocities approach the limit c for high frequencies; there is a low-frequency cutoff when $V_{l,m}$ is infinite (and $U_{l,m}=0$),

$$\frac{1}{c^2}\omega_{l,m}^2 = \left(\frac{\pi l}{a}\right)^2 + \left(\frac{\pi m}{b}\right)^2, \quad \text{cutoff.} \tag{9}$$

A certain l, m wave can be generated by an oblique plane wave falling into the wave guide (Fig. 2). Let θ, $\varphi_x\varphi_y$ be the angles of the incident beam with z, x, and y. Let us take

$$\pm\cos\varphi_x = -\frac{c}{\omega}\frac{\pi l}{a} = \frac{l\lambda}{2a}, \quad \pm\cos\varphi_y = -\frac{c}{\omega}\frac{\pi m}{b} = \frac{m\lambda}{2b}, \tag{10}$$

$$\lambda = \frac{c}{\nu} = \frac{2\pi c}{\omega},$$

and

$$\cos^2\theta + \cos^2\varphi_x + \cos^2\varphi_y = 1,$$

hence,

$$\frac{\omega^2}{c^2}\sin^2\theta = \left(\frac{\pi l}{a}\right)^2 + \left(\frac{\pi m}{b}\right)^2. \tag{11}$$

The incident plane wave is reflected back and forth on the walls of the guide. Each reflection changes the sign of either $\cos\varphi_x$ or $\cos\varphi_y$. The superposition of all these reflected waves yields a resultant wave (6) and comparison of (8) with (11) gives the phase velocity V_{lm} in the z direction.

$$V_{lm} = \frac{c}{\cos\theta} > c. \tag{12}$$

Cut-off conditions obtain for $\theta \doteq \pi/2$.

High l, m terms correspond to very oblique rays, and these terms are needed when a good definition of point M is required. A convenient superposition of proper waves (6) can be used to represent the light spreading out from a light source located at M. We shall use a finite number of terms in the sum (6)

$$0 \leqslant l \leqslant l_M \quad 0 \leqslant m \leqslant m_M, \tag{13}$$

and we select a frequency well above cut-off conditions. This yields a concentration of light around M, with an extension

$$\Delta x \sim \frac{a}{l_M}, \quad \Delta y \sim \frac{b}{m_M}. \tag{14}$$

The errors on x (and y) correspond to the shortest distances between nodes in the x (or y) direction. This problem is completely discussed in the Appendix.

We cannot separate the variables z and t since our waves are propagating freely in the z direction, and we assume the wave guide to be properly terminated to avoid reflections at both ends. In order to define the instant of observation, we use short pulses of duration Δt, and these short pulses can be repeated periodically at intervals τ (duration of one single observation). It is easy to show (see Appendix) that these conditions yield, for each wave, a number of degrees of freedom

$$n \approx \frac{\tau}{2\Delta t}. \tag{15}$$

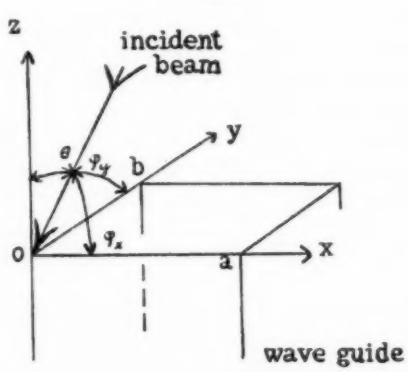


FIG. 2.

Altogether, each proper wave (l,m) in the wave guide has n degrees of freedom and we have a total of

$$nl_M m_M = \frac{\tau}{2\Delta t}\frac{a}{\Delta x}\frac{b}{\Delta y}, \tag{16}$$

degrees of freedom in a beam capable of specifying the position M within the errors, Δx, Δy, Δt.

III. DISCUSSION

Let us first assume a low accuracy, enabling us to use a low frequency ($h\nu < kT$). Each degree of freedom takes at a temperature T an energy kT, hence, a total thermal energy

$$E_T = nl_M m_M kT. \tag{17}$$

We certainly must use in the beam an energy E at least equal to AE_T, where A is the coefficient defined in the previous paper (see reference 1, Sec. 4, p. 1156), if we want to distinguish our light source M from the background of thermal radiation. When an observation is made, the light scattered from M is absorbed at I in a photoelectric cell, or on a photoelectric plate. This results in an entropy increase

$$\Delta S = \frac{E}{T} = nl_M m_M kA \quad A > 1. \tag{18}$$

Comparing Eqs. (5) and (18) we have

$$\begin{aligned} \Delta I &= k \ln F, \\ \Delta S &= \tfrac{1}{2} A k F, \end{aligned} \tag{19}$$

with

$$F = \frac{\tau}{\Delta t}\frac{a}{\Delta x}\frac{b}{\Delta y}.$$

The negentropy principle of observation is satisfied since

$$\Delta S - \Delta I > 0, \quad E = \frac{\Delta I}{\Delta S} < 1, \tag{20}$$

and we end up with conditions very similar to those previously obtained in other problems (see reference 1, Sec. 9, Eqs. 68, 69).

When high accuracy is required, one must use light of high frequency ($h\nu > kT$). This is the case of usual optical instruments. Thermal background is very low at high frequencies but we certainly must use at least one quantum $h\nu$ on each degree of freedom of the beam, hence the energy E dissipated and the entropy increase ΔS for one observation:

$$\begin{aligned} E &= nl_M m_M h\nu, \\ \Delta S &= knl_M m_M \frac{h\nu}{kT}. \end{aligned} \tag{21}$$

The coefficient A of Eq. (18) is replaced by $h\nu/kT$, which may be a very large number. Let us assume low accuracy in y and t but high accuracy on the x measurement. This means that l_M is large, while m_M and n are moderately large (Eqs. 14, 15). The frequency ν is obtained from Eq. (9):

$$\nu = \frac{c}{2}\left[\left(\frac{l_M}{a}\right)^2 + \left(\frac{m_M}{b}\right)^2\right]^{\frac{1}{2}} \approx \frac{cl_M}{2a}, \quad l_M \approx \frac{a}{\Delta x}, \tag{22}$$

hence,

$$\frac{h\nu}{kT} \sim \frac{hc}{2kT\Delta x} = \frac{1}{2}\frac{\Delta x_T}{\Delta x}, \quad \Delta x_T = \frac{hc}{kT} \approx \frac{1.44}{T}, \tag{23}$$

where Δx_T is the characteristic length at temperature T, as defined in reference (1), Eq. (54). Here again, as in the preceding paper (see reference 1), we conclude that the measurement of very small distances requires very high entropy increases, and that this entropy cost may become prohibitive at nuclear dimensions.

IV. UNCERTAINTY RELATIONS IN THE MICROSCOPE EXPERIMENT

The microscope experiment is usually discussed, in connection with uncertainty relations, under the assumption that one single photon can be used to obtain information about the position of the particle. This seems to be an oversimplified assumption. A photon materializing at a point I may come from anywhere and does not necessarily indicate a scattering particle at M. The preceding discussion clearly shows that a large number of photons is required to define unambiguously a focal image at I. Let us re-examine the problem of uncertainty relations on this example.

The conditions of observation are the same as on Fig. 1. We now want to measure x over an interval a, with an error Δx, but we are no more interested in y or b or Δy. In the proper waves (6) we pay attention to the factor $\cos(\pi lx/a)$ and to the l values, while ignoring m and the $\cos(\pi my/b)$ factor, whenever this is feasible. For instance we may choose $b > a$, thus making the ground wave $l=0$ and $m=1$, since these values yield the lower cut-off frequency in Eq. (9). This wave shall represent the incident beam, coming up through the wave guide and striking on particle M:

$$\textit{Incident beam} \quad l=0, \quad m=1. \tag{24}$$

This incident beam represents the superposition of two plane waves with

$$\cos\varphi_y = \pm\frac{\lambda}{2b},$$

according to Eq. (10). We can also think of m taking positive or negative values and assume $m = \pm 1$.

Light scattered from particle M may contain proper waves with all possible l, m values. We can get rid of most of the m terms by using in L a cylindrical lens of length b. This device cancels out all terms but for

$m=0$. We thus describe the scattering process as a transition,

$$\begin{cases}\text{in } l\cdots\text{from 0 to } l,\\ \text{in } m\cdots\text{from 1 to 0.}\end{cases} \tag{25}$$

But, in order to account for the $\pm$ terms in the Eq. (10) we had better consider transitions

$$\begin{cases}\text{in } l\cdots\quad 0\rightarrow\pm l,\\ \text{in } m\cdots\pm 1\rightarrow 0.\end{cases}$$

A change δl during scattering of a photon results in an impulse δp_x to the M particle,

$$\delta p_x=\frac{h}{\lambda}\delta(\cos\varphi_x)=\frac{h}{2a}l \tag{26}$$

according to (10). The $\pm$ sign is now contained in l, and the value l may run from $-l_M$ to -1 or from $+1$ to $+l_M$. The value $l=0$ is excluded because l and m cannot both be zero. This takes account of condition (14). Condition (15) specifies that we must use n terms on each wave in order to define t with an error Δt. Since impulses δp_x can be positive or negative, we compute the sum of squares and call it $(\Delta p_x)^2$,

$$(\Delta p_x)^2=\sum_n\sum_l(\delta p_x)^2$$

$$=n\left(\frac{h}{2a}\right)^2 4[1+4+9+\cdots l_M{}^2], \tag{27}$$

$$\Delta p_x=\left(\frac{h}{a}\right)\left[\frac{n}{3}l_M(l_M+1)(l_M+\tfrac{1}{2})\right]^{\frac{1}{2}}.$$

We have 4 cases in (25), with $\pm\, l$ and ± 1, hence the factor 4 in Eq. (27). We now use the values (14) for l_M and (15) for n, hence,

$$\Delta p_x\Delta x=h\left[\frac{\tau}{6\Delta t}\left(1+\frac{\Delta x}{a}\right)\left(\frac{a}{\Delta x}+\frac{1}{2}\right)\right]^{\frac{1}{2}}. \tag{28}$$

This is our improved uncertainty relation, which takes into account the time definition and the range of variation of the variable x, in the microscope experiment.

TABLE I.

$\frac{\tau}{\Delta t}$	$\frac{a}{\Delta x}$	$\left[\frac{\tau}{6\Delta t}\left(1+\frac{\Delta x}{a}\right)\left(\frac{a}{\Delta x}+\frac{1}{2}\right)\right]^{\frac{1}{2}}$
2	1	1
2	2	$\frac{1}{2}\sqrt{5}=1.12$
2	3	$\frac{1}{2}\sqrt{14}=1.25$
3	1	$\sqrt{1.5}=1.22$
3	2	$\frac{1}{2}\sqrt{\frac{15}{2}}=1.37$
large values		$\left[\frac{\tau}{6\Delta t}\frac{a}{\Delta x}\right]^{\frac{1}{2}}$

The value of τ may range from $2\Delta t$ up to infinity. The value $\tau=\Delta t$ is out of order since it does not give any information. The parameter a may range from Δx to infinity. Let us tabulate a few simple cases (see Table I). When accuracy is very low, and most of the definition is due to screens and obstacles ($a=\Delta x$) the usual uncertainty obtains. But we may be interested in cases where the time definition and the measurement of position be mostly obtained through the optical system, and this leads to much higher uncertainty. Let us take for instance

$$\frac{\tau}{\Delta t}=6000,\quad \frac{a}{\Delta x}=1000,$$
$$\Delta x\Delta p_x=h10^3. \tag{29}$$

We thus come to the conclusion, that the uncertainty may become very large when we want to measure the position of a "free" particle: free to move over a large distance during a long time, without being disturbed by material screens or shutters or any mechanical device.

V. MEASUREMENT OF MOMENTUM

In the microscope observation we assumed that we knew the initial velocity of the particle, and we tried to measure the position. The results were not very favorable. Let us try the reverse operation. We start by measuring the position x at time t with great accuracy. This gives a large unknown impulse to the particle (Eq. 28) but it does not matter, since we did not know the previous velocity v_x anyhow. Then we can measure v_x and p_x by Doppler effect.[4] The usual argument gives

$$\Delta x\Delta p_x\sim h. \tag{30}$$

The standard discussion is based on the assumption that one single photon $h\nu'$ may suffice for a good measurement of the frequency ν' in the Doppler effect, provided a long wave packet of total duration θ can be used

$$\Delta\nu'\cdot\theta\sim 1. \tag{31}$$

We now want to investigate this problem more carefully and to take advantage of the remarks based on the theory of information. We intend to show that one single photon is not enough and we shall discuss how the number of photons required in the experiment is related with exact specifications on the experimental procedure. This will lead us to a modified uncertainty relation

$$\Delta x\Delta p_x\sim nh,\quad n\gtrsim 1, \tag{32}$$

where n depends upon the conditions of observation. This practical limit may be lowered to the previous limit (30) when conditions are selected in such a way as to make n practically unity.

The discussion about information taught us that we must always define the range of variation and error on

[4] See, for instance; D. Bohm, *Quantum Theory* (Prentice-Hall, Inc., New York, 1951), p. 105.

the quantities to be measured. In our present problem, we specify:

$$\begin{cases}\text{time } t \text{ of observation}, & \text{error } \Delta t=\theta, \text{ range } \tau,\\ \text{velocity } v_x \text{ of the particle}, & \text{error } \Delta v_x, \text{ range } u.\end{cases} \tag{33}$$

As in previous examples, we introduce the range τ in time by assuming that flashes of duration θ are repeated at intervals τ. These light signals can be obtained by superimposing n different frequencies (See Appendix, Eq. A.14),

$$n=\frac{\tau}{2\Delta t}=\frac{\tau}{2\theta}. \tag{34}$$

These frequencies extend over a frequency range $\delta\nu=1/\theta$. We also specify in (33) the range u of velocities over which we want to operate. We shall design a system capable of measuring velocities between v_0 and v_0+u with an error Δv. This means using m resonators tuned on m different Doppler frequencies ν', each resonator being able to receive a frequency band $\Delta\nu'$ corresponding to Δv.

$$m=\frac{u}{\Delta v}, \quad \Delta\nu'=\frac{\nu}{c}\Delta v, \tag{35}$$

ν, average initial frequency.

The measurement is possible, provided the original frequency range $\delta\nu$ of Eq. (34) be smaller than the $\Delta\nu'$ introduced in (35). The problem of an experiment requiring m simultaneous observations was discussed in Sec. 4 of a previous paper (reference 1) and it was shown that one should select on each resonator an energy limit E_l

$$E_l=A_m kT. \tag{36}$$

$E<E_l$ means that the resonator is considered unexcited, $E>E_l$ corresponds to an excited resonator. Calling r the reliability factor of the experiment, and assuming high reliability, we computed

$$A_m\approx\ln(r\cdot m), \quad r\gg 1, \quad m\gg 1. \tag{37}$$

These conditions can be satisfied by using a single quantum $h\nu'$ provided

$$h\nu'\geqslant A_m kT.$$

The energy E_l will correspond to a single quantum on each frequency ν' if we take

$$h\nu'=E_l\approx A_m kT, \tag{38}$$

and we have n different frequencies ν', covering a band $\delta\nu$ (Eq. 34) in order to define the instant of time of the measurement. Finally, the energy dissipated is $nh\nu'$ and the entropy increase

$$\Delta S=\frac{nh\nu'}{T}\approx nA_m k=kn\ln(rm). \tag{39}$$

The Information obtained is

$$\Delta I=k\ln\frac{\Delta t}{\tau}\cdot\frac{\Delta v}{u}=k\ln\frac{n\cdot m}{2}, \tag{40}$$

and the generalized second principle is satisfied

$$\Delta(S-I)=k\left[(n-1)\ln m+n\ln r-\ln\frac{n}{2}\right]>0. \tag{41}$$

The numerical factor n may be small, but is always larger than 2. We assumed high reliability and high accuracy ($r\gg 1, m\gg 1$) and ΔS is much larger than ΔI.

Since we are using n quanta, we obtain a recoil n times greater than in the usual discussion and the uncertainty relation is given by Eq. (32).

As an example, let us assume

$$r=e^{10}, \quad m=e^{10}, \quad A_m=20,$$

according to (37). Hence, by Eq. (38)

$$h\nu'\approx 20kT.$$

If the experiment is made at room temperature, this condition indicates frequencies near the visible range. Information theory and quantum condition do not come into any contradiction but their conditions complete each other. For applications to field theory we shall use Eq. (30) which represents the lower limit of Eq. (32).

VI. UNCERTAINTY IN FIELD MEASUREMENTS

Uncertainty relations for a field theory were discussed by N. Bohr and L. Rosenfeld[5] for the case of electromagnetic fields. A probe of volume V, charge ϵ is used during a time τ. The momentum p_x is measured at t, and takes another value p_x' at $t+\tau$. The average field $\bar{E}_x$ over V during τ is

$$\bar{E}_x=\frac{p_x'-p_x}{\epsilon\tau}. \tag{42}$$

The position x of the probe is measured with an error Δx and results in an error Δp_x on p_x' and p_x. We contend that we can use, in this case, the relation (30). We first measure x, then p_x at time t (Sec. 5). Then we measure p_x' at $t+\tau$ and finally x'. The last determination throws off the particle with a large unknown momentum, but we do not care any more. The errors on p_x give an error on the field

$$\Delta\bar{E}_x\approx\frac{h}{\epsilon\tau\Delta x}\approx\frac{h}{e\tau Z\Delta x}, \quad \epsilon=Ze, \quad e \text{ electron charge}. \tag{43}$$

Observations under a microscope enable us to measure the field from its action on a free particle, with no

[5] N. Bohr and L. Rosenfeld, Kgl. Danske Videnskab Selskab, Mat.-fys. Medd. XII, paper 8 (1933); W. Heitler, *The Quantum Theory of Radiation* (Oxford University Press, London, 1944), Chapter II, p. 76.

physical obstacles perturbing this field. In this formula, we may take for Δx the diameter of the particle, and $Z\Delta x$ is a characteristic of the particle used as a probe. It may range from 10^{-13} for an electron up to 10^{-10} for a heavy nucleus ($Z=96$, $\Delta x \sim 10^{-12}$). In order to define a time interval τ we need a beam of light with frequencies ranging from 0 to ν_M (see Appendix, Eq. A.14),

$$\nu_M = \frac{1}{2\tau}, \tag{44}$$

hence,

$$\epsilon \Delta \bar{E}_x \Delta x \approx 2h\nu_M. \tag{45}$$

This formula has a simple interpretation, since $\epsilon \bar{E}_x \Delta x$ is the work done by the field $\bar{E}_x$ on a distance Δx. This work cannot be measured with great accuracy because we may have emission or absorption of quanta $h\nu_M$ coming from the beam of light used to define the time interval τ.

Altogether, our discussion does not modify the fundamental formula of Bohr and Rosenfeld, but it emphasizes some difficulties in regard to uncertainty relations. The usual uncertainty relations correspond to an absolute theoretical limit, but they cannot be reached in many experimental devices, where a much higher uncertainty is obtained. This is especially the case when the position of a practically free particle is to be measured (see Sec. 4). The discussion confirms the results of a previous paper,[1] where the special difficulty of measuring very small distances was emphasized.

APPENDIX

We discussed in Sec. 2 the propagation of waves in a rectangular wave guide, and we stated the conditions (14) giving the accuracy with which a focus can be defined, when a finite number of proper waves is used.

Let us examine the situation more carefully. The pipe's boundaries act as perfect mirrors, reflecting the pipe's cross section indefinitely (Fig. 3). A focus F_{00} within the pipe has images at $F_{-1,0}$, $F_{0,-1}$, $F_{-1,1}$ in the next rectangles, and so on indefinitely. We must use a superposition of waves (6) representing this situation in the focal plane $z=0$. In this plane, all components are in phase, in the expansion (6). For a different z value, these same components get out of phase, because of their different V_{lm} velocities and the beam spreads out. Considering only the focal plane $z=0$, we have

$$H_{z=0} = \sum_{l,m} A_{lm} \cos l\phi \cos m\psi \cos \omega t$$

$$\text{with} \quad \phi = \frac{\pi x}{a}, \quad \psi = \frac{\pi y}{b}; \tag{A.1}$$

and we may split the double Fourier series in two separate series in x and y, writing

$$\begin{aligned} H_{z=0} &= X(x)Y(y)\cos\omega t, \\ X(x) &= \sum_l X_l \cos l\phi, \quad 1 \le l \le l_M, \\ Y(y) &= \sum_m Y_m \cos m\psi, \quad 1 \le m \le m_M, \\ A_{lm} &= X_l Y_m, \end{aligned} \tag{A.2}$$

we do not include $l=0$ or $m=0$ in order to be sure not to introduce the 0; 0 forbidden terms.

The problem is similar along x or y. Let us consider the x coordinate only. We want to obtain a focus F_{00} and all its images. This means sharp maxima of $X(x)$ at points

$$\begin{aligned} &x = x_0 + 2pa, \quad x = -x_0 + 2pa, \\ &p \text{ positive or negative integer.} \end{aligned} \tag{A.3}$$

We shall build a series with a finite number of terms representing such a situation with a given accuracy. Let us start from Lagrange's identity

$$\begin{aligned} &\cos\phi + \cos 2\phi \cdots + \cos l_M\phi \\ &= \sum_1^{l_M} \cos l\phi = \tfrac{1}{2}\left(\frac{\sin(l_M+\tfrac{1}{2})\phi}{\sin\tfrac{1}{2}\phi} - 1\right) \\ &= \frac{\sin\dfrac{l_M}{2}\phi \cos\dfrac{l_M+1}{2}\phi}{\sin\dfrac{\phi}{2}}. \end{aligned} \tag{A.4}$$

We now take

$$1 \le l \le l_M, \quad X_l = \cos l\phi_0, \quad \phi_0 = \frac{\pi x_0}{a}, \tag{A.5}$$

hence,

$$\begin{aligned} X(x) &= \sum_1^{l_M} \cos l\phi \cos l\phi_0 \\ &= \tfrac{1}{2}\sum_1^{l_M} [\cos l(\phi-\phi_0) + \cos l(\phi+\phi_0)], \end{aligned}$$

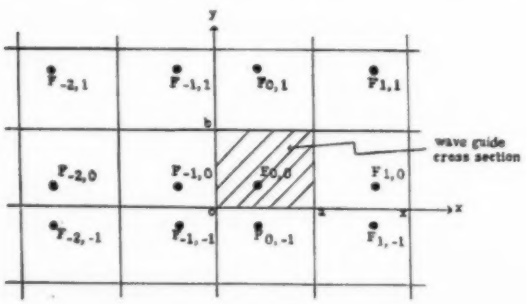


FIG. 3.

and we use the formula (A.4)

$$X(x)=\frac{\sin\frac{l_M}{2}(\phi-\phi_0)\cos\frac{l_M+1}{2}(\phi-\phi_0)}{2\sin\frac{\phi-\phi_0}{2}}+\frac{\sin\frac{l_M}{2}(\phi+\phi_0)\cos\frac{l_M+1}{2}(\phi+\phi_0)}{2\sin\frac{\phi+\phi_0}{2}}. \quad (A.6)$$

The first term yields sharp maxima at

$$\phi=\phi_0+2p\pi,$$

when its denominator is zero. The second term gives maxima at

$$\phi=-\phi_0+2p\pi,$$

and this gives the complete system of focus and images (A.3).

The series of coefficients (A.5) gives exactly the conditions required. Let us consider the original focus itself and its neighborhood

$$\begin{aligned} &x=x_0+\delta x,\\ &\phi=\phi_0+\delta\phi \quad \delta x=\frac{a}{\pi}\delta\phi. \end{aligned} \quad (A.7)$$

The second term in Eq. (A.6) is negligible. The first term has a maximum value $l_M/2$ at the focus ($\delta\varphi=0$) and obtains its first zeros at

$$\delta\phi=\pm\frac{\pi}{l_M+1},$$

when the cosine term in the numerator of the first term is zero. The width of the maximum can be practically defined as $\frac{1}{2}$ the distance between the first zeros to the right and to the left

$$\begin{aligned} &\Delta\phi=\tfrac{1}{2}(2\delta\phi)=\frac{\pi}{l_M+1},\\ &\Delta x=\frac{a}{l_M+1}\approx\frac{a}{l_M}. \end{aligned} \quad (A.8)$$

When l_M is large, this condition results in our previous relation (14).

Similar considerations apply to the y coordinate. It is easy to verify that these conditions correspond exactly to the usual definition of the separating power.

Let us consider a square aperture and take equal limits along x and y:

$$a=b=r, \quad \Delta x=\Delta y, \quad l_M=m_M=M.$$

The radius of the focus is

$$\Delta r=(\Delta x^2+\Delta y^2)^{\frac{1}{2}}=r\left(\frac{1}{l_M{}^2}+\frac{1}{m_M{}^2}\right)^{\frac{1}{2}}=\sqrt{2}\frac{r}{M}. \quad (A.9)$$

The limiting angle θ_0 is obtained from Eq. (11),

$$\frac{4r^2}{c^2}\sin^2\theta_0=\left(\frac{l_M}{a}\right)^2+\left(\frac{m_M}{b}\right)^2=2\left(\frac{M}{r}\right)^2, \quad (A.10)$$

hence,

$$\Delta r=\frac{c}{\nu\sin\theta_0}=\frac{\lambda}{\sin\theta_0}, \quad (A.11)$$

which is the usual formula for the separating power as a function of wave length λ and angular aperture θ_0.

Let us now examine condition (15). It results directly from Lagrange's identify (A.4),

$$f(\theta)=\tfrac{1}{2}+\sum_1^n\cos n\theta=\frac{\sin(n+\frac{1}{2})\theta}{2\sin\frac{1}{2}\theta}, \quad \theta=\frac{2\pi t}{\tau}. \quad (A.12)$$

This superposition of terms represents a succession of pulses at

$$\begin{aligned} \sin\tfrac{1}{2}\theta=0, \quad &\theta=0,\ 2\pi,\ 4\pi,\\ &t=0,\ \tau,\ 2\tau. \end{aligned}$$

The duration of the pulse occurring at $t=0$ can be defined as $\frac{1}{2}$ the distance between the two zeros on both sides of the maximum

$$(n+\tfrac{1}{2})\theta=\pm\pi, \quad (2n+1)t=\pm\tau,$$

hence,

$$\Delta t=\frac{\tau}{2n+1}, \quad (A.13)$$

which reduces to Eq. (15) when n is large.

The function $f(\theta)$ represents a string of pulses, Δt repeated at intervals τ. It contains frequencies ranging from 0 to ν_M,

$$\nu_M=\frac{n}{\tau}\approx\frac{1}{2\Delta t}. \quad (A.14)$$

When $f(\theta)$ is used as a modulation on a carrier of frequency ν, the pulse system contains frequencies from $\nu-\nu_M$ up to $\nu+\nu_M$ hence, a frequency band

$$\Delta\nu=2\nu_M=\frac{1}{\Delta t}. \quad (A.15)$$

This corresponds to Eq. (31).

Decompositions of Inorganic Specimens During Observation in the Electron Microscope*

ROBERT B. FISCHER
Department of Chemistry, Indiana University, Bloomington, Indiana
(Received November 18, 1953)

Several inorganic substances have been examined by electron microscope and selected area electron diffraction techniques. Large crystals of sodium chloride and potassium chloride break up into many smaller crystals of the same substances. Ammonium chloride evaporates, and nitron nitrate appears merely to melt. Reduction to metal was found for silver chloride, silver nitrate, lead carbonate, and cupric oxide. Only morphological changes were observed on bombardment of silver sulfate, lead oxide, lead chloride, and gold chloride. Hydrous ferric oxide and aluminum oxide were converted to crystalline forms which could not be identified by electron diffraction.

INTRODUCTION

SPECIMENS in electron microscopy and electron diffraction are exposed to combined conditions of evacuated surroundings and electron bombardment during observation. The possibility that specimens may undergo decomposition or other alterations during mounting and observation has been recognized since the beginnings of electron microscopy.

Burton, Sennet, and Ellis[1] reported observations of several crystalline substances as the specimens were left in a fairly intense electron beam for several minutes. Crystals of sodium chloride, sodium chlorate, potassium bromide, and potassium iodide were all opaque to the electron beam when first observed, but became transparent upon continued exposure. Originally opaque silver chloride particles became partially transparent. Watson[2] reported further observations on several other substances, including tetracopper calcium oxychloride, which became granular during observation, and tungsten oxide, which changed from platelets to a lace-like material.

The nature of specimen decompositions is of utmost importance in electron microscopy and in electron diffraction work, and consideration must be given to the possibility of decomposition or other alteration of every type of specimen. The present study was undertaken, with the aid of selected area diffraction techniques, in an attempt to provide further insight into the nature of these decompositions.

EXPERIMENTAL

Specimens were mounted on parlodion or formvar covered nickel grids. Some specimen particles were applied to the mounting supports by evaporation, either in air or in vacuum, of a solution of the test substance. Other specimens were mounted dry by lightly touching a film-covered grid to dry, powdered material. An RCA type EMU electron microscope, equipped with both selected area and standard electron diffraction attachments, was employed.

Electron diffraction patterns, whenever of the powder type, were identified by comparison of experimental data with data listed in the standard A.S.T.M. card files. In order to obtain reproducible selected area diffraction patterns, the following steps were adopted: regular objective lens focus controls were set for sharpest focus of the microscopic image prior to photographing each selected area diffraction pattern; diffraction focus control set to position corresponding to sharpest center spot of each pattern; all patterns were measured lengthwise on the film.

OBSERVATIONS AND DISCUSSION

1. Sodium Chloride and Potassium Chloride

Sodium chloride crystals were found, as reported earlier, to decompose during observation from opaque particles to transparent envelopes outlining the original

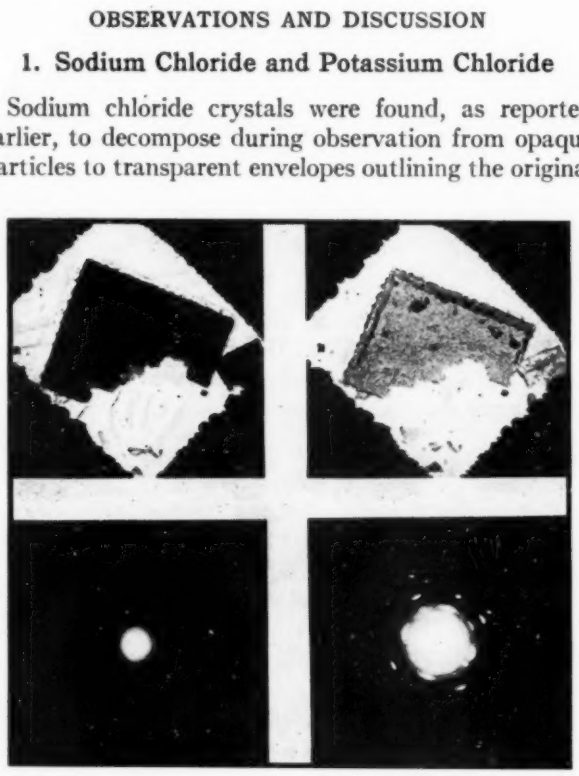


FIG. 1. Micrographs of sodium chloride and corresponding selected area electron diffraction patterns (left: before decomposition in the beam), (right: after decomposition in the beam).

* Publication number 607 from the Chemistry Department of Indiana University, Bloomington, Indiana.

[1] Burton, Sennet, and Ellis, Nature **160**, 565 (1947).

[2] J. H. L. Watson, J. Appl. Phys. **19**, 713 (1948).

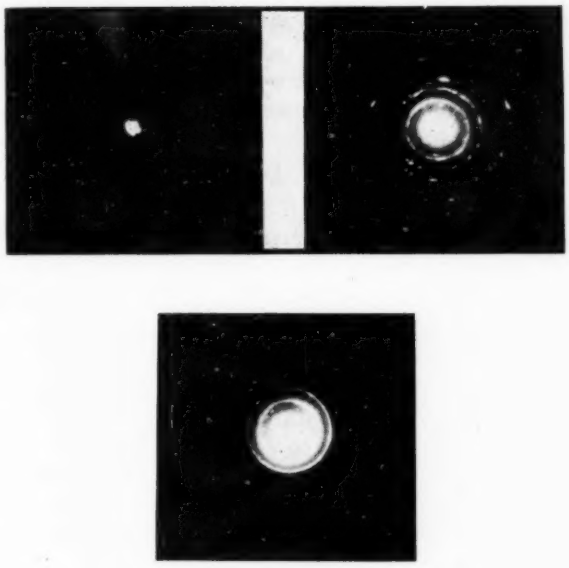


FIG. 2. Selected area electron diffraction patterns of potassium chloride, before, during, and after decomposition in the beam.

particles. Selected area diffraction patterns of the initial and final crystals, Fig. 1, reveal that the relatively few crystals of sodium chloride initially present are replaced by a large number of tinier crystals, still of sodium chloride. Thus, the sodium chloride does not melt and evaporate, but rather it breaks up into many tinier crystals of the same material. It is not as yet possible to ascertain the significance of the remaining envelope; it likely includes many tiny sodium chloride crystals, thus contributing to the diffraction pattern. Preparations of sodium chloride crystals which are appreciably larger or smaller than those illustrated in Fig. 1, do not undergo this type of decomposition upon initial observation.

The decomposition of potassium chloride crystals proceeds in much the same manner as that of sodium chloride crystals. The progression from one crystal to many crystals within a given area of the specimen mount is illustrated by the selected area diffraction patterns of Fig. 2; the same substance, potassium chloride, is revealed by all measurable patterns. The intensity of electron bombardment of the specimen in the regular diffraction position of the EMU instrument was found inadequate to cause appreciable decomposition, while the more intense bombardment in the upper position did accomplish ready decomposition. This fact suggests that, although heat does not appear to be the only factor of significance, it is very much involved in the observed decompositions.

2. Ammonium Chloride and Nitron Nitrate

Ammonium chloride particles were observed to undergo ready decomposition, often starting as a sputtering and often continuing until seemingly nothing remained. Selected area diffraction patterns in most instances revealed nothing crystalline, either initially or finally; on a few occasions faint spots and/or rings appeared briefly during the decomposition. Nitron nitrate precipitates, consisting of well-defined needles, appear merely to melt and spread out on the supporting film. These decompositions of ammonium chloride and of nitron nitrate may be largely or even entirely thermal melting and/or evaporation.

3. Silver Chloride

A precipitate of silver chloride was prepared by mixing solutions of sodium chloride and silver nitrate, and removing excess electrolyte by ion diffusion through a parlodion membrane. The precipitate consisted initially of chain-like aggregates of opaque particles which changed during observation to smaller opaque globules within transparent envelopes outlining the initial particles, Fig. 3. In addition, there was sputtering of material onto nearby areas of the parlodion film. Initial selected area diffraction patterns consisted of rings which were spotty, due to relatively few particles available for contribution to the pattern, and which were of silver chloride. Final diffraction patterns were of free silver. Thus the silver ion of silver chloride is reduced during observation to free silver. It was noted that clear areas of the parlodion film removed several microns from the nearest specimen particles, yielded excellent selected area diffraction patterns of metallic silver after near-by silver chloride particles had been reduced, even though the film remained microscopically clear. Thus, free silver must have sputtered out quite some distance from decomposing particles, and it must deposit an extremely thin layer which is transparent to the electron beam.

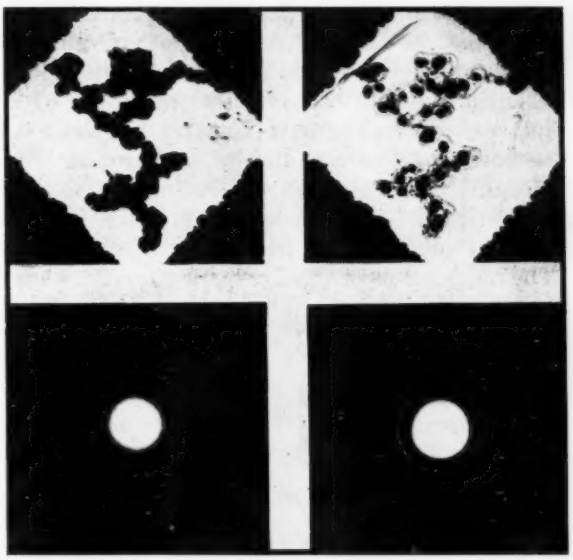


FIG. 3. Micrographs of silver chloride and corresponding selected area electron diffraction patterns (left: before decomposition in the beam), (right: after decomposition in the beam).

FIG. 4. Aluminum oxide after decomposition in the beam.

4. Silver Nitrate and Sulfate; Cupric Oxide; Gold Chloride; Lead Carbonate, Oxide and Chloride

These substances were examined in an effort to see if other reduction reactions could be accomplished by the electron beam. Reductions to free metal were found to occur with silver nitrate, cupric oxide, and lead carbonate. No reductions to the metallic state were found to occur with silver sulfate, gold chloride, lead oxide, (PbO) or lead chloride. Some morphological decomposition resulted from electron bombardment with each one of these materials. It is suggested that the occurrence of these reduction reactions is of significance, and that this type of reaction may be fairly common in electron microscopy. The reaction is highly selective, as indicated by the fact that silver nitrate does reduce and silver sufate does not, and the fact that lead carbonate does while lead oxide and lead chloride do not. The anion must be disposed of somehow during each reduction, yet copper oxide reduces and lead oxide does not. The decomposition of lead carbonate does not consist of the two steps of loss of carbon dioxide and then reduction to metallic lead, since lead oxide itself does not reduce to free metal.

The mechanisms of these reduction reactions are not clear. It is doubtful that all are photochemical reductions, and the electron beam may itself serve directly as a reducing agent. The reducing action of bombardment with electrons is of interest in considering decompositions of organic specimens as well as of inorganic ones. For example, the suggestion of König[3] that collodion type films become graphite films during observation would necessitate the presence of a reducing agent.

5. Hydrous Ferric Oxide and Aluminum Oxide

A gelatinous precipitate of hydrous ferric oxide was prepared by mixing solutions of ferric chloride and ammonium hydroxide and removing excess electrolyte by ion diffusion through a parlodion film. Microscopic observation of this precipitate revealed many irregular masses of particles, some of which underwent shrinkage and other morphological alteration during electron bombardment. Several different crystalline forms were indicated by selected area diffraction patterns, and nearly all areas of the mounts were crystalline. Thus, some dehydration or other change of the hydrous material must have occured even before initial observation.

An aluminum oxide preparation was found to decompose markedly during observation, and a variety of diffraction patterns were obtained from the decomposition products. None of the diffraction patterns of this preparation or of the ferric oxide preparation could be identified with certainty as any of the many known crystalline forms of aluminum oxide, iron oxide, aluminum metal or iron metal, but combinations of two or more forms may have been obtained each time. The aluminum oxide frequently decomposed to leave a sputtered residue along with opaque x-shaped particles. The micrograph, Fig. 4, reveals an appearance similar to a shadow-cast specimen, although this material does not appear useful for an "internal shadowing" technique.

ACKNOWLEDGMENT

The author wishes to acknowledge with appreciation the technical assistance of Carl J. Wust and Richard Childness.

[3] König, H., Nachr. Wiss. Gottingen, Math.-physik Klasse, Math.-physik.-chem. Abt. **1946**, 24 [Abstracted in Chem. Abstracts **43**, 7324 (1949)].

Arcing at Electrical Contacts on Closure. Part V. The Cathode Mechanism of Extremely Short Arcs

PAUL KISLIUK
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received November 18, 1953)

The available experimental data on the short arcs which occur on closure of electrical contacts with potential differences less than the minimum sparking potential are shown to be consistent with the hypothesis that the arcs are initiated and sustained by field emission electrons. Positive ions then carry less than 7 percent of the current at the cathode. The primary function of the ions is to supply sufficient positive space charge to maintain the high cathode field, but the action of the discrete ions near the cathode in creating a temporary pass in the potential barrier may result in the attainment of sufficient electron emission to sustain the arc at values of the cathode field somewhat lower than that required by the Fowler-Nordheim equation for the same current density.

It is shown that the supply of volatile impurities on the electrode surfaces is critical in determining the breakdown field and the minimum arc current.

IN a series of papers[1–4] Germer, Haworth, and Smith have investigated the short arcs which occur at separations of the order of 2000A when electrical contacts are closed at moderate potential differences (~50 volts). These arcs are not initiated by the usual mechanism in air because their breakdown voltage (as low as 30 volts) is far below any reported values of the "minimum" sparking potential.[5,6] The present paper uses the measurements of these earlier papers and certain further results to develop the postulate that the arcs are both initiated and maintained by field emission electrons.

The breakdown voltages of short gaps have been investigated by several workers by older methods which did not reveal the ensuing arcs,[7–10] as well as more recently by oscilloscopic methods.[11,12] These studies show that the breakdown occurs at a roughly constant apparent field, dependent on the electrode material and surface condition, of the order of three million volts per centimeter for reasonably clean contacts in air. Roughness of the surface creates local fields several times as great. The apparent breakdown fields are considerably higher for carefully polished gold surfaces,[10] for electrodes immediately after heavy arcing,[13] and for well outgassed electrodes in high vacuum.[14] One is led to conclude from the high breakdown fields that field emission is the means of obtaining the initial electrons.

A discharge sustained by field emission has been proposed for the mercury arc and for arcs of other low boiling metals.[15–18] To maintain the high cathode fields required for such a mechanism the current density must be even higher than those recently observed of the order of 10^6 amps/cm^2,[19] and this mechanism is considered questionable by some authors.[20] For the extremely short arcs under consideration, however, the current densities are very high. A lower limit of 3×10^7 amps/cm^2 is obtained from the arc currents and the area of the erosion pit of a single arc.[1] This, together with the high initiating field, is sufficient to remove the objections to the field emission mechanism of sustaining and initiating short arcs.

The observed arc voltages are comprehensible in terms of the field emission hypothesis since the cathode fall in potential must be somewhat greater for field emission cathodes than for other means of emission.[21] Most of the electrons are emitted from the cathode region with a potential energy less than that at the cathode surface by the work function ϕ (Fig. 1), and those that suffer no collisions arrive at the anode with an energy $V-\phi$, where V is the arc voltage. In order that the ionization cross section be greater than zero in the region near the anode this energy must be greater than the ionization potential V_i by a small amount Δ. Thus we obtain

$$V-\phi=V_i+\Delta; \quad V=V_i+\phi+\Delta. \tag{1}$$

In Table I the measured arc voltages are compared with the sum $V_i+\phi$.

With the possible exception of carbon, which may produce an arc of a different type, the agreement is striking considering the uncertainty in some of the

[1] L. H. Germer and F. E. Haworth, J. Appl. Phys. **20**, 1085 (1949).
[2] L. H. Germer, J. Appl. Phys. **22**, 955 (1951).
[3] L. H. Germer, J. Appl. Phys. **22**, 1133 (1951).
[4] L. H. Germer and J. L. Smith, J. Appl. Phys. **23**, 553 (1952).
[5] L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (John Wiley and Sons Inc., New York, 1939), p. 414.
[6] J. D. Cobine, *Gaseous Conductors* (McGraw-Hill Book Company Inc., New York, 1949), p. 165.
[7] R. F. Earhart, Phil. Mag. **1**, 147 (1901); **16**, 48 (1908).
[8] P. E. Shaw, Proc. Roy. Soc. (London) **73**, 337 (1904).
[9] C. Kinsley, Phil. Mag. **9**, 692 (1905).
[10] G. L. Pearson, Phys. Rev. **56**, 471 (1939).
[11] Reference 1, pp. 1090–2.
[12] Reference 4, pp. 557–9.
[13] M. M. Atalla (private communication).
[14] P. Kisliuk (to be published).
[15] I. Langmuir, Gen Elec. Rev. **26**, 731 (1923); Science **58**, 2010 (1923).
[16] K. T. Compton, J. Am. Instr. Elec. Engrs. **46**, 868 (1927).
[17] S. S. Mackeown, Phys. Rev. **34**, 611 (1934).
[18] T. Wasserab, Z. Physik **130**, 311 (1951).
[19] K. D. Froome, Proc. Phys. Soc. (London) **63**, 377 (1950).
[20] L. B. Loeb, see reference 5, pp. 629–636.
[21] R. C. Mason, Phys. Rev. **38**, 427 (1931).

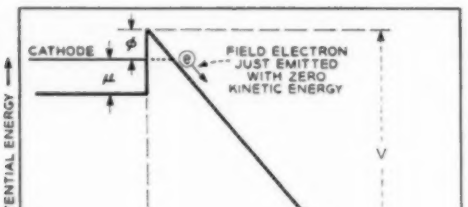


FIG. 1. Energy diagram for a typical field emission electron.

work functions, especially under uncertain surface conditions due to contamination. The minimum voltages obtained from published values of the Ayrton constants[22] from measurement on longer arcs, are generally not in agreement with the values in Table I, indicating either the presence of an anode fall, which may not exist for short arcs, or an entirely different process of cathode emission.

The mechanism outlined above requires that many electrons cross the gap without collisions. This is consistent with the small gap, which is less than one mean free path at atmospheric pressure. Under such circumstances the bulk of the arc energy is dissipated at the anode surface, which is quickly raised to the boiling point, creating a dense vapor at the surface to supply the arc with ions. In the case of longer vacuum arcs, which are also characterized by small values of the pressure-distance product, the anode is not so effective in supplying vapor to the arc, and evaporation from the cathode becomes of importance.[23,24]

For short arcs the first molecules to leave the anode must be evaporated by the heating due to field emission

TABLE I.

	Arc voltage (L.H.G.)[a]	(Gaulrapp)[b]	(Rossbach and Seeliger)[c]	$V_i+\phi$[d]
Cu	12.	12.6	12.6–15.6	7.7+4.5=12.2
Ag	11.	12.3		7.5+4.3=11.8
Pd	14.	—		8.3+4.8=13.1
Pt	15.	15.3		8.9+5.3=14.2
Au	12.	12.6		9.2+4.4=13.6
C	20–30	—	12.1–12.3	11.2+4.4=15.6
Zn	11.5	10.9		9.4+3.7=13.1
W	—	15.2	11.8–12.7	8.1+4.5=12.6

[a] See reference 2.
[b] K. Gaulrapp, Ann. Physik 25, 705 (1936).
[c] P. Rossbach and R. Seeliger, Z. Physik 116, 68 (1940).
[d] Values of ϕ are from Michaelson, J. Appl. Phys. 21, 536 (1950), and those of V_i from the Handbook of Chem. and Physics (1947).

[22] R. Seeliger, *Einfuhrung in die Physik der Gasentladung* (J. Barth, Leipzig, 1934), p. 374.
[23] W. P. Dyke and J. K. Trolan, Phys. Rev. 89, 799 (1953).
[24] R. Holm, *Electric Contacts* (H. Gebers, Stockholm, 1946), p. 269.

electrons. In high vacuum with well outgassed electrodes, these first molecules are necessarily metal atoms, if arcs thus initiated are possible under these circumstances. In nearly all practical examples, however, the layers of adsorbed gas will supply these initial molecules, and metal vapor appears only when the current density is increased by the presence of positive space charge in the gap.[25]

The presence of atmospheric gases is thus expected to influence the breakdown only through adsorbed layers, and moderate vacuum should have little or no effect on short arc initiation. This is indeed the case. From atmospheric pressure down to 10^{-3} mm Hg, breakdown fields for gaps less than about 10^{-4} cm remain in the region of a few million volts/cm, as long as the voltage thus obtained is below the sparking potential in air. An experimental curve illustrating this observation is given in Fig. 2.

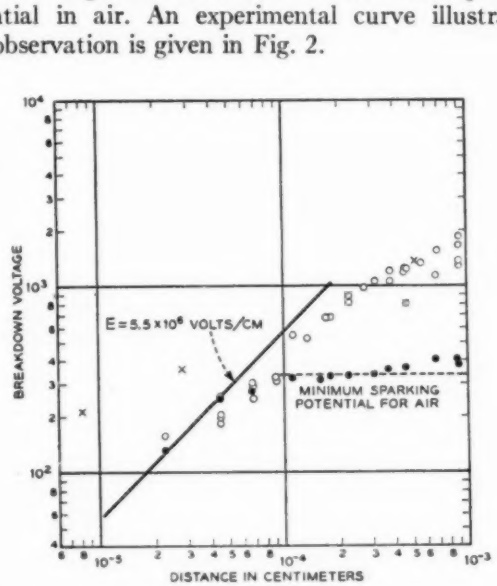


FIG. 2. Breakdown voltages for palladium contacts. ● .... in room air. ○ .... in moderate vacuum. × .... in high vacuum ($\sim 10^{-8}$ mm Hg).

In vacua on the order of 10^{-8} mm Hg with outgassed electrodes, the breakdown fields may be as high as 25 million volts/cm for extremely small gaps, but the breakdown voltage no longer follows a constant field relationship. This is illustrated in Fig. 2, but will be discussed in a subsequent paper in which the apparatus used in obtaining these experimental points will also be described.

Layers of foreign matter on the electrodes are of importance in determining the minimum arc current as well as the breakdown field. For clean contacts in the open air the minimum arc current is of the order of one ampere,[26,27] while, for surfaces contaminated by the products of certain organic vapors charred by previous arcing, the arc may be maintained at 0.1 ampere

[25] See reference 24, p. 270.
[26] See reference 24, p. 264.
[27] See reference 2, p. 963.

or even less.[27] In high vacuum, on the other hand, the arc may fail while many amperes are flowing.[28]

The explanation of the above probably lies in the fact that the supply of ionizable molecules in the gap is critical. For sufficiently contaminated and porous surfaces small currents can maintain the local high pressure because the loosely adhering material is easily evaporated, and the arc travels along the surface from one dirty spot to another as the intense heat and bombardment cleans off the surface. For well outgassed metal surfaces in vacuum much higher currents are necessary to maintain the required density of metal atoms in the gap.

The available measurements of short arc characteristics may be used to gain a further insight into their mechanism. Mackeown[17] has given a first integral of Poisson's equation on the assumption that the cathode fall space is less than one ionic mean free path. This may be written

$$E_c^2 = 7.57\times10^5 V_c^{\frac{1}{2}} I\{x[(1845W)^{\frac{1}{2}}+1]-1\} \qquad (2)$$

Where E_c is the cathode field; V_c is the cathode fall in potential; I is the total current density at the cathode; x is the fraction of the current carried by positive ions at the cathode; and W is the molecular weight of the ions, all in practical units.

We can estimate the fraction of the current at the cathode carried by ions in the short arcs under consideration by substituting in Eq. (2) the following values:

(1) For V_c, the total arc voltage. Table I.

(2) For I, a lower limit of 3×10^7 amps/cm^2 as explained earlier.

(3) For E_c, 5×10^7 volts/cm, which is obtained from the Fowler-Nordheim equation on the assumption of a field emission mechanism with current densities of the order of magnitude used for I and a work function of 4.5 volts.

One obtains an upper limit for x of about 0.07, a value less than that associated with other types of arcs. Other effects, to be considered subsequently, which increase the electron emission and therefore permit the attainment of the observed current densities at somewhat lower fields, result in a still lower value of x. Wasserab[18] has given graphical solutions of Eq. (2), assuming a field emission mechanism and taking $V_c=10$. He finds that in mercury arcs the electron current must greatly exceed the ion current at the cathode, and that the current densities in the cathode spot must exceed the largest values measured experimentally. Thus field emission arcs are efficient in their utilization of ions to produce electrons, and may be expected to prevail under conditions of low pressure-distance product, when ions are in short supply. This is the case when the gap is very small as in low voltage arcs on closure and for longer arcs under vacuum conditions.

Field emission arcs may occur also under circumstances unfavorable to thermionic emission at the cathode, as when the cathode is a low boiling metal, and under certain conditions, during the formative period of arcs of the high boiling metals.[29] In the latter case the presence of adsorbed gas is of great importance,[30] just as it is for short arcs on closure.

Since more than 90 percent of the current at the cathode is carried by electrons, many of which arrive at the anode without collision, a large part of the arc energy is dissipated at the anode by electron bombardment. This is almost wholly utilized in melting and boiling metal.[31] Some of the boiled metal must condense on the cathode, resulting in the transfer of a certain amount of heat and anode material. However, if the fraction of the current at the cathode were as much as 10 percent ionic, the entire measured transfer of 9×10^{-14} cm^3/erg[32] could be accounted for as ions which remain on the cathode after neutralization. This is interpreted to mean that the cathode current is in fact less than 10 percent ionic, in which case there can also be considerable condensation of neutral atoms.

A computation of the heat delivered to the cathode by both ions and condensing atoms boiled from the anode only accounts for $\frac{1}{3}$ of the measured value of 42 percent of the arc energy.[33] Thus, there must be other means of returning the kinetic energy gained by the electrons in the field to the cathode.

The particle density built up by the evaporated metal atoms is so great that after the initial stages of the arc the mean free path becomes less than the gap distance and some of the energy gained by the gas molecules in the space through electron impact may be returned to the cathode, as is the case in longer arcs. One can estimate a lower limit for the pressure and particle density from the measured amount of metal transferred, the volume contained between the anode and cathode erosion pits, and the arc duration, under the assumption that metal is boiled from the anode and condensed on both electrodes. This yields a lower limit of about 55 atmospheres at the boiling point of platinum, or a particle density at least six times that at N.T.P. The gap is then somewhat greater than one mean free path, but the arc may burn around the periphery of the high pressure region. Since the computed pressure is an average over the pit area and the arc duration, the maximum pressure is higher, if the pressure is not constant over these intervals of time and area.

---

[28] R. Holm, see reference 24, pp. 265, 270.

[29] O. Becken and R. Seeliger, Ann. phys. **24**, 609 (1935).

[30] O. Becken and R. Sommermeyer, Z. Physik. **102**, 551 (1936).

[31] See reference 1, p. 1088.

[32] See reference 1, p. 1086. (The value of 4×10^{-14} cm^3/erg given in this reference is based on the energy stored in the capacitor. The figure quoted above is obtained from this and is based on the arc energy.)

[33] L. H. Germer, Bell System Tech. J. **30**, 933 (1951).

From the knowledge of the ion current density j^+ and the arc voltage one can compute the ion charge density ρ^+, in the cathode fall space $[\rho^+ \approx j^+/(eV/m)^{\frac{1}{2}}]$. This is on the order of 10^{19} ions/cm^3. To obtain this high figure, the particle density must be much greater than that at N.T.P., in keeping with the calculation above.

Germer and Smith[4] have shown that the steady-state arc voltage can be established within, at most, a few millimicroseconds after the first appreciable drop in voltage. This is of the order of fifty ion transit times. In order that the arc conditions be built up within the observed time, the number of ions and electrons must be increased by a factor Z for each ion transit time. The smallest current which would discharge the capacity of the oscilloscope plates and lead wires at a rate visible in the oscillogram is about 10^{-3} ampere, while the minimum arc current is about 1 ampere, so that $Z_{\min}{}^{50} = 10^3$, whence $Z_{\min} = 1.15$. It is probable that the actual value of Z is considerably higher than this lower limit. Furthermore, since the gap is only about $\frac{1}{3}$ of a mean free path, and the efficiency of ionization by elecrons in the established arc is at most 0.07 ion/electron, the number of ions produced per electron is less than $\frac{1}{3}$. Therefore the number of electrons produced by each ion must be several times Z. Thus, there must be more than four electrons produced by each ion, probably many more, during the formative period of the arc.

This is of the order of magnitude predicted by Newton[34] for the number of additional electrons liberated by the local fields of a single ion as it approaches a metal cathode in an average field of about 10^7 volts/cm, a field for which the Fowler-Nordheim equation yields a current density far too small to sustain the arc. This calculation, however, is beset by several uncertainties as to the validity of the assumptions needed to obtain a solution, in particular the choice of a distance of nearest approach of the ions before they are neutralized or reflected. In the sense of this mechanism the electron emission might not be due solely to the average high field at the cathode, but might be augmented by the succession of ions each of which temporarily creates a "pass" in the potential barrier. Even in this case the space charge due to ions too distant from the cathode to participate in this effect must be invoked to maintain cathode fields sufficiently high to obtain the necessary electron emission. This mechanism, too, might be termed a field emission arc, since it is due to a barrier penetration effect and fairly high fields are necessary, and the considerations of this paper are then changed only by substituting a somewhat lower value of E_c in Eq. (2), thereby obtaining an even lower value of x which is consistent with the consideration of the transfer of matter in the arc. Similarly such effects as photoelectric emission, which may supply electrons or enhancement of the field by ions on thin insulating films,[35,36] with relatively high efficiency because of the high field, will serve to reduce the percentage of ionic current at the cathode.

In conclusion, it has been shown that the postulate that short arcs on closure are initiated and sustained by field emission electrons, perhaps with some enhancement due to the discrete nature of the ions, is in harmony with the existing experimental data. The ion current at the cathode is then at most a few percent of the total. It is shown that the initiating field is strongly dependent on surface contamination, which, in addition to affecting the emission of electrons from the cathode, supplies the initial atoms for ionization when the contacts are in the open air or under moderate vacuum.

The author would like to thank Drs. L. H. Germer, W. S. Boyle, and M. M. Atalla for helpful discussions during the course of this work.

[34] R. R. Newton, Phys. Rev. **73**, 1122 (1948).

[35] F. E. Haworth, Phys. Rev. **80**, 223 (1950).

[36] F. Llewellyn Jones, Proc. Inst. Elec. Engrs. (London) **100**, part I, 169 (1953).

Shadowing Technique for Electron Microscopy—A Possible Substitute for the High-Vacuum Evaporation Technique

ALVIN E. BILLS AND ROBERT LEFKER
Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey
(Received December 28, 1953)

A method of shadowing samples directly within the electron microscope has been developed in this Laboratory. This method, which uses thallium chloride immersed in distilled water as the shadowing material, is much less time consuming than is the usual method that involves high-vacuum metal evaporation; the results obtained are not of the same high quality as those of the usual method, but the time saving often justifies the method. Polystyrene Latex (batch No. 580) is introduced on the specimen screen, to provide an internal calibration standard; this standard makes it possible to measure the angle of shadowing.

INTRODUCTION

THE shadow-casting technique in electron microscopy yields very valuable information, in many cases, concerning the particle size and morphology of various materials. The use of the high-vacuum evaporator requires the expenditure of much time, even though high-quality results may not be required. While shadowing the specimen *in situ* in the electron microscope will not give high-quality results, the method here described is a great timesaver as compared with the high-vacuum evaporator technique.

This method of shadowing was first observed during an attempt to prepare TlCl for use as a standard in electron diffraction work. It was noticed that when a particle of TlCl was brought into the field of view and subjected to bombardment by an electron beam of relatively high intensity, the TlCl evaporated and redeposited on other parts of the screen,[1,2] leaving behind the skeletal remains of the original particle. It was observed that some of the particles of TlCl, which evidently did not receive the full force of the electron beam and thus did not evaporate, caused shadows to form in a direction away from the position of the evaporated particle[3] (Fig. 1).

EXPERIMENTAL PROCEDURE

A small portion of thallium chloride is placed on one edge of a Formvar-coated screen which has had a sample deposited on its surface. This is most easily done by taking a portion of the TlCl powder and immersing it in water; the chloride is sparingly soluble—0.32 g/100 ml—in cool water. A small portion of the immersed chloride is lifted from the water and placed with care on the screen; a toothpick of the "Worlds Fair" type is here used for this procedure. The dampness of the powder provides a loose bond with the film on the screen.

The screen is then placed in the sample holder, sample side up; the holder is placed in the microscope. The column is evacuated, and the electron beam is obtained by the usual procedure.

As the illuminating intensity is increased, the beam impinging on the particles of TlCl causes them to melt or vaporize and spread over the surface of the screen as heretofore described and as illustrated in Fig. 1. The particles of specimen on the screen, of course, impede the progress of the migrating minutae of shadowing material. This produces the effect of casting shadows alee of the direction in which the chloride particles are traveling; the shadow formation is similar to that in the high-vacuum metal evaporator technique of shadowing.

Selected area diffraction may be used to advantage in locating a desirable field of shadowed specimen. Since TlCl renders an exceedingly well-defined diffraction pattern, the operator may check a promising-looking field for TlCl minutae by merely switching to selected area diffraction and noting the resulting diffraction pattern. Portions where the diffraction pattern is beginning to weaken indicate that the coating of shadowing material is very heavy or that it is very light. Particles of shadowing material of the smallest grain size are found in the areas in which the coating is light.

FIG. 1. Micrograph of a field close to the source of TlCl—*in situ* method of shadowing (mag. 10 500×).

[1] Burton, Sennett, and Ellis, Nature **160**, 565 (1947).
[2] J. H. L. Watson, J. Appl. Phys. **19**, 713 (1948).
[3] Burton, Sennett, and Ellis, Nature **160**, 565 (1947).

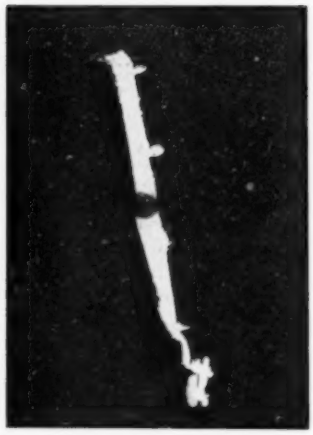


FIG. 2. Micrograph of single particle of an asbestos-like synthetic material shadowed with TlCl—the *in situ* method (7200×).

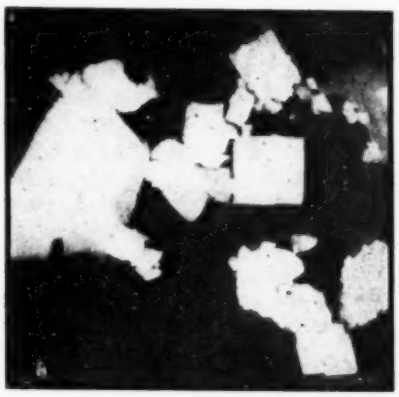


FIG. 4. Micrograph of cristobalite shadowed with chromium in high-vacuum evaporator for comparison with Fig. 5 (7200×).

RESULTS

This technique was tried on various samples—among them cristobalite, polystyrene Latex, and asbestos-like synthetic material. Micrographs of these samples are shown in Figs. 2 through 5. It may be seen that the angles of shadowing vary from micrograph to micrograph, and it may be seen that the shadows are all cast from low angles. The portion of the specimen nearest the original particles of shadowing TlCl is densely covered as shown in Fig. 1. Viewable areas of the screen located at intermediate distances from the source of evaporated material contain more desirable fields (Figs. 2, 3, 5) than do areas close to, or at considerable distances from, the source. Figure 2 shows that within a viewable field, the more distant the specimen is from the shadowing source, the smaller are the shadowing minutae and the less distinct are the shadows The shadow angle decreases as the distance between the specimen and TlCl source increases. Figures 4 and 5 show comparisons between the high-vacuum evaporator shadowing method and the shadowing *in situ* method.

DISCUSSION

The shadow angle cannot be determined beforehand, nor has this Laboratory yet devised a means of controlling the angle. The angle of shadowing, however, may be determined in this way: before the introduction of the TlCl, a drop of polystyrene Latex (batch No. 580) solution is placed on the specimen screen and allowed to dry; the length of the shadows cast by the polystyrene-Latex spheres may then be measured on the photographic plate, and the shadowing angle calculated.

Care in handling TlCl is necessary; the salt is poisonous. Good practice requires the thorough cleaning of the specimen screen holder and the lower part of the specimen pole piece after the completion of shadowing. Investigation of other shadowing substances is continuing, in the hope that TlCl may be replaced by a

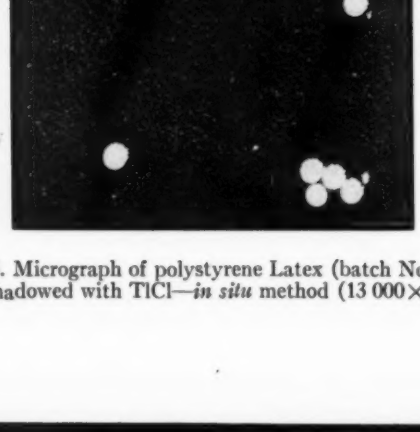


FIG. 3. Micrograph of polystyrene Latex (batch No. 580) shadowed with TlCl—*in situ* method (13 000×).

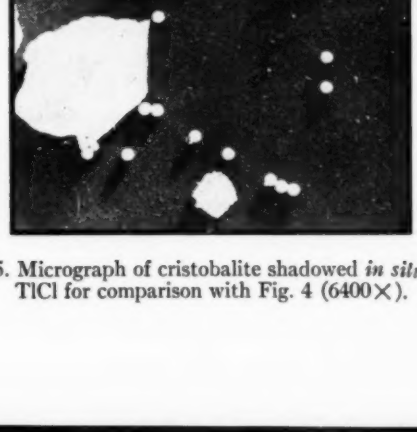


FIG. 5. Micrograph of cristobalite shadowed *in situ* with TlCl for comparison with Fig. 4 (6400×).

material which is nonpoisonous and which emit minutae of smaller grain size. Materials which have been tried experimentally are cobaltous chloride and sodium chloride. Cobaltous chloride yields minutae of small grain size; the salt is unsuitable, however, because it does not shadow enough of the screen area. Sodium chloride seems to have definite possibilities; it yields minutae of relatively small grain size, and it covers the screen quite well when subjected to bombardment.

Rough determinations of particle morphologies may be made by the new method in much less time than that needed to make similar determinations by the high-vacuum evaporator methods. The saving in time has proved to be significant, in the work done by the authors.

ACKNOWLEDGMENTS

The authors are indebted to William F. Nye for making possible this study and to Ellis L. Manning for his help in preparing the manuscript.

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 7 JULY, 1954

Measurement of the Conductivity of a Jet Flame*

FRED P. ADLER
Hughes Research and Development Laboratories, Culver City, California
(Received November 30, 1953)

A method for the measurement of the complex conductivity of a rocket exhaust flame is described. Typical results are shown for an acid-aniline rocket motor and a frequency of 200 mc/sec. The importance of sodium impurities is pointed out and illustrated experimentally.

I. INTRODUCTION

THE extremely hot exhaust flame of a rocket motor may contain a sufficiently high concentration of free electrons to produce appreciable attenuation of electromagnetic waves propagating through the jet. Such attenuation can cause serious practical problems in the communication systems of jet-propelled crafts. The study of the effect of such flames upon electromagnetic radiation is therefore of considerable interest.

The basic quantity which determines the effect of an ionized medium on radiation is its complex conductivity. From a knowledge of this conductivity the insertion loss of a flame may be calculated, including both the effect of the impedance mismatch in the boundary region and the attenuation within the flame. A method for the measurement of the conductivity of a jet exhaust is described in this paper, as well as some typical results which have been obtained.

II. EXPERIMENTAL TECHNIQUE AND APPARATUS

The method of measurement used is basically identical with that described in an earlier paper.[1] As shown in Fig. 1, the apparatus consists essentially of a frequency-modulated oscillator exciting a cylindrical resonant cavity, and of a receiver which in conjunction with an oscilloscope, gives a measure of the power output from the cavity. The ionized gas is introduced along the cavity axis.

The cavity used in the tests performed at the Jet Propulsion Laboratory is shown in Figs. 2 and 3. It was designed to resonate in the fundamental TM_{010} mode at 200 mc/sec, corresponding to an inside diameter of 115 cm. To provide the requisite ruggedness and heat capacity, ½-inch steel was used as the material. The interior surfaces were copper-plated to reduce eddy-current losses. Circular openings at the centers of the walls served as intake and exit holes for the rocket flame. To couple into and out of the cavity, loops of copper tubing connected to the center conductor of the coaxial cable were placed on the side walls so as to link with the magnetic field. About 80 feet of lossy cable were used in order to attenuate reflections and eliminate standing waves.

The oscillator was tuned to the resonant frequency of the cavity and frequency-modulated over about a 40 kc/sec range at a rate of 60 cps. With the amplified

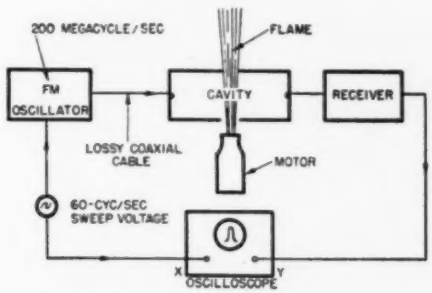


FIG. 1. Experimental arrangement.

* This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. DA-04-495-Ord 18, sponsored by the Department of the Army, Ordnance Corps. The experimental work described was performed in spring of 1950, but was cleared only recently for open publication.

[1] F. P. Adler, J. Appl. Phys. **20**, 1125 (1949).

FIG. 2. Çavity.

output from the cavity applied to the vertically deflecting plates of the oscilloscope the resonance curve of the cavity was traced out, the maximum amplitude being proportional to the output field strength. As shown below, measurement of the change in this amplitude and of the shift in resonant frequency due to the flame is sufficient to calculate both real and imaginary components of the conductivity. By photographing the trace before and after turning on the motor, both of the quantities were determined.

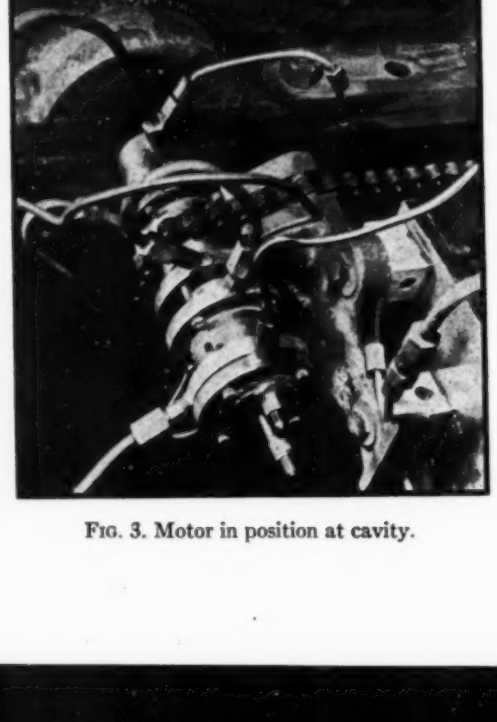


FIG. 3. Motor in position at cavity.

An experimental 50 pound-thrust, acid aniline motor, shown in Fig. 3, was used in these experiments. Red fuming nitric acid served as the oxidizer; the fuel consisted of a mixture of aniline and furfuryl alcohol.

III. THEORETICAL RELATIONS

A. Theoretical Conductivity Expression

In deriving a rigorous expression for the conductivity from kinetic theory it is found necessary to make one of two simplifying assumptions: the electronic mean free path is independent of the electronic velocities and temperature; or the mean free time τ between collisions does not change with temperature. The meaning of these two hypotheses is discussed elsewhere.[2] The formula based on the first assumption[3] is not too readily amenable to computations. The second assumption, however, leads to the following simple expression for the conductivity σ:

$$\sigma=\frac{ne^2\tau}{m}\frac{1-j\omega\tau}{1+\omega^2\tau^2}=\sigma_r-j\sigma_i. \tag{1}$$

Here e and m are the charge and mass of an electron; n denotes the electron density; and ω is the radian frequency of the electromagnetic wave. Although the actual situation in the present experiment is probably intermediate between the two assumptions mentioned, Eq. (1) will be used here because of its greater simplicity.

It is seen that if both σ_r and σ_i are known, the average electron collision frequency ν_e can be found immediately since

$$\nu_e=\frac{1}{\tau}=\omega\frac{\sigma_r}{\sigma_i}. \tag{2}$$

If this expression is substituted in Eq. (1), the electron concentration may be solved for in terms of the measured values of σ_r and σ_i

$$n=\frac{m\omega}{e^2}\left(\frac{\sigma_r}{\sigma_i}+\frac{\sigma_i}{\sigma_r}\right)\sigma_r. \tag{3}$$

B. Calculation of the Conductivity from Experimental Measurements

The pertinent relations as given below follow directly from the theory developed in the previous paper.[1] The real part of the conductivity is found from

$$\sigma_r=1.69\times10^{-3}\frac{a}{b^2}\frac{1}{Q}(r-1), \tag{4}$$

where a is the cavity radius, b the radius of the flame which is considered to be in the shape of a cylinder coaxial with the cavity, Q is the familiar quality factor of the cavity in the absence of the jet flame, and r

[2] F. P. Adler and H. Margenau, Phys. Rev. **79**, 971 (1950).
[3] H. Margenau, Phys. Rev. **69**, 508 (1946).

denotes the ratio of the field strengths at the receiver without and with the jet flame present. Thus, r is merely the ratio of amplitudes of the oscilloscope traces without and with the rocket-motor firing. Putting in the appropriate measured values, $Q=12\,700$, $b\approx 1.6$ inches, yields

$$\sigma_r=4.8\times10^{-5}(r-1) \text{ mhos/meter.} \quad (5)$$

The imaginary component of the conductivity was shown to be related to the frequency shift Δf (in cps) which it produces by the following expression:

$$\sigma_i=4\pi\epsilon_0\left(\frac{a}{b}\right)^2\frac{J_1^2(ka)}{J_0^2(kb)+J_1^2(kb)}\Delta f. \quad (6)$$

Here ϵ_0 is the dielectric constant of free space, and $k=2\pi/\lambda$, λ being the free-space wavelength (1.50 meters). Substituting the proper numerical values for the present case results in

$$\sigma_i=2.37\times10^{-8}\Delta f \text{ mhos/meter.} \quad (7)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The main results of this investigation are shown in Table I. It is seen that during smooth running of the motor the conductivity is quite low, a representative value being 0.2×10^{-4} mhos/meter. In a homogeneous ionized medium the corresponding attenuation of a radio wave would amount to only 0.033 db/meter. It is seen that there is a slight increase in σ_r with flame temperature. Somewhat higher values of conductivity were observed during the initial transient while the motor is starting up. These values may be correlated with the bursts and minor explosions occurring during this period, as illustrated in Fig. 4 which shows a relatively smooth start.

TABLE I. Measured values of conductivity (σ_r).

Oxidizer-fuel ratio	1.65	2.20	1.96
Approximate theoretical temperature (°F)	2100	2600	3000
Average field-strength ratio (r)	0.735	0.710	0.705
Average $\sigma_r\times10^5$ (mhos/meter)	1.72	1.97	2.02
Initial-transient $\sigma_r\times10^5$ (mhos/meter)	···	4.25	4.45

The determination of the imaginary conductivity component proved to be more difficult. This measurement depends on observing the increase in the resonant frequency of the cavity due to the presence of the flame. Unfortunately the initial temperature rise of the cavity amounted to about 1°F/sec, corresponding to a resonant-frequency decrease of almost 1 kc/sec/sec. Hence frequency-shift data have to be taken during the first few fractions of a second after firing. Thus data were obtained only if the motor started up very smoothly. Such a start was obtained only for the test at a mixture ratio of 1.65 for which a shift of 1.44 kc/sec was recorded

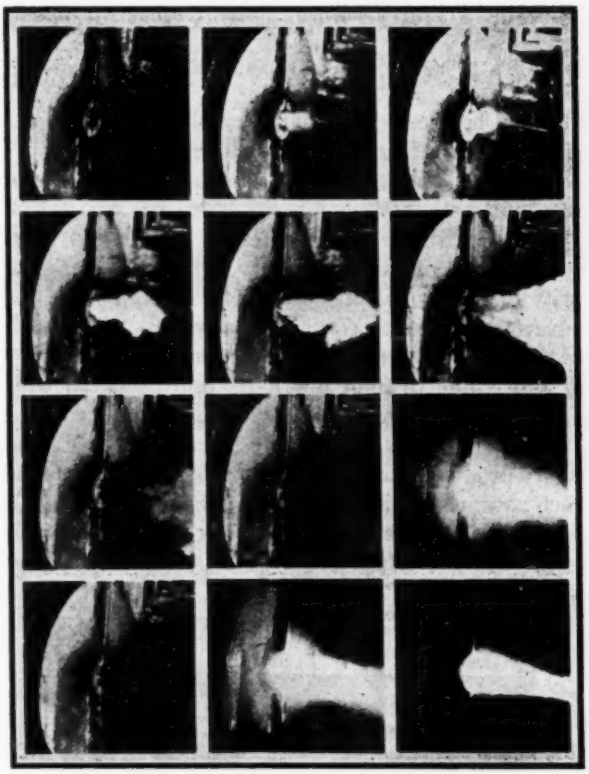


FIG. 4. Starting sequence (16 frames/sec) of 50-lb-thrust motor.

corresponding to $\sigma_i=0.34\times10^{-4}$ mhos/meter. Using Eq. (2), the mean collision frequency can now be found

$$\nu_c=\omega\frac{\sigma_r}{\sigma_i}=6.5\times10^8 \text{ collisions/sec.}$$

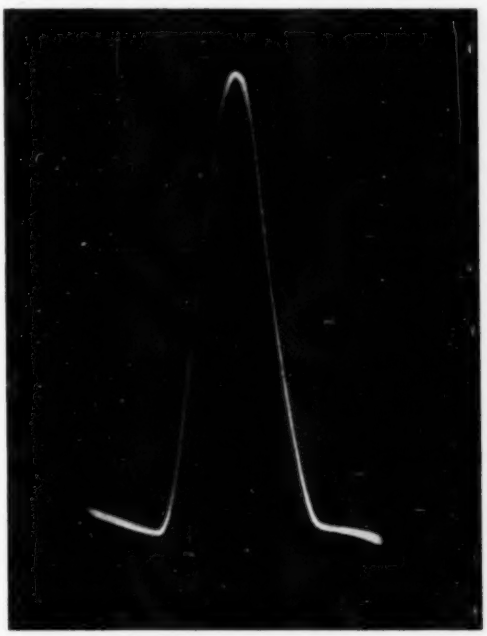


FIG. 5a. Reference resonance curve of empty cavity.

Equation (3), furthermore, gives the electron density.

$$n=4.5\times10^{10}(\sigma_r/\sigma_i+\sigma_i/\sigma_r)\sigma_r=1.9\times10^{6}\text{ electrons/cc.}$$

It should be mentioned that the value of σ_i just given may be quite inaccurate; nevertheless, the values of ν_c and n shown are believed to represent useful order-of-magnitude estimates.

The surprisingly low values of conductivity and electron concentration are very probably due to the purity of the propellants used; specifically, to the freedom from impurities containing any salts of the alkali metals. The importance of these metals, particularly of sodium, is due to their low ionization potential which makes for a relatively high degree of ionization at jet-flame temperatures. A rough estimate of the electron density produced can be obtained from Saha's equation[4] which assumes equilibrium conditions and neglects any secondary processes in the flame. For sodium at a representative flame temperature of 1800°K this equation gives

$$n_{\text{Na}}\approx2.2\times10^{10}(fp)^{\frac{1}{2}}\text{ electrons/cm}^3,$$

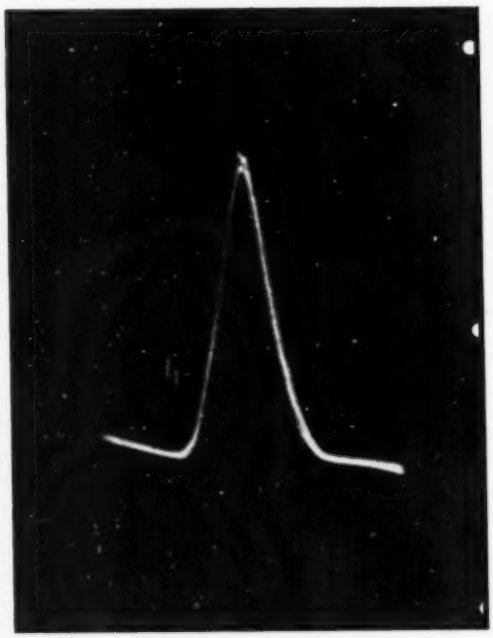


FIG. 5b. Effect of flame on resonance curve when no impurities are present.

[4] M. N. Saha, Phil. Mag. **40**, 472 (1920).

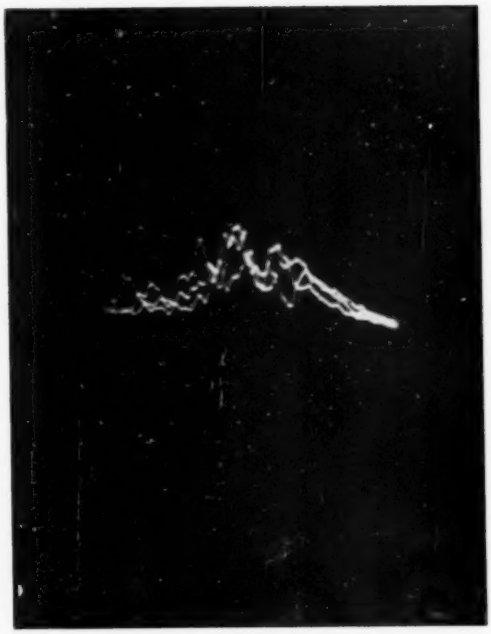


FIG. 5c. Effect of flame on resonance curve when sodium impurities are added.

where f is the weight fraction of the Na and p is the total pressure in atmospheres. It is seen that an impurity amounting to only 10 ppm of sodium would produce a higher electron density than the one calculated from the present measurements. It must then be concluded that essentially no sodium impurities were present in the propellants used.

As a purely qualitative check on the importance of sodium, a very small fraction of sodium lactate was added to the acid in the final test run. The results shown in Fig. 5 are quite striking. The signal is now strongly attenuated and appears very unsteady and noisy.

V. ACKNOWLEDGMENT

The author is indebted to Professor W. H. Pickering for suggesting this study and extending the use of the facilities of the Jet Propulsion Laboratory. The writer writer also wishes to thank Mr. C. Hylkema for his support of this program, Mr. J. Wedel and Mr. D. Hughes for their work on the rf equipment, and Mr. D. Thomas for engineering the rocket-motor firings.

Contact Charging between Nonconductors and Metal*

JOHN W. PETERSON†
University of California, Berkeley, California
(Received December 22, 1953)

An investigation has been made of the charging of fused quartz and borosilicate glass spheres rolling on a clean nickel surface. The results indicate that the process of acquiring charge depends only upon the nature and surface condition of the materials in contact and is independent of the pressure of an atmosphere of dry N_2. The total charge, however, is limited by gaseous discharge to the metal and therefore does depend upon the pressure. Such gaseous discharge has been verified by the observation of individual discharge pulses. The rate of acquiring charge and maximum charge in vacuum has been found to depend upon rolling speed, both reaching a maximum at an intermediate speed, and to decrease with increasing surface conductivity. Since both fused quartz and borosilicate glass charge negatively in contact with clean nickel, it is inferred that both have higher effective work functions than nickel. The surface conductivity of borosilicate glass has been found to be three of four orders of magnitude higher than that of quartz. Therefore, since the charging of the two materials is of the same order of magnitude, the effective work function of the glass appears to be considerably higher than that of fused quartz.

PREVIOUSLY reported work[1] upon the charging of borosilicate glass spheres in contact wlth nickel confirmed the variation of charging with pressure, first observed by Debeau.[2] It also indicated that charging depends upon speed of rolling and surface conductivity, and that a good vacuum is necessary in order to obtain reproducible results. Indicated improvements over the first work included elimination of sharp impacts, substitution of one sphere for a group of spheres to eliminate interactions among them, and a continuous record of the charging process. The apparatus used in the present work was designed to incorporate these refinements, and to make possible quantitative study of all the factors known to be important to contact charging.

Figures 1 and 2 show schematic views of the apparatus. An accurately ground six-millimeter sphere[3]

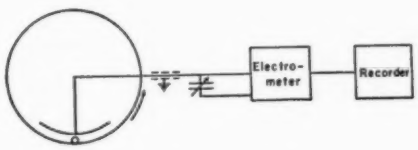


FIG. 1. Schematic diagram of apparatus.

of the nonconductor to be investigated rolls freely in the bottom of a nickel cylinder revolving about a horizontal axis. The cylinder is mounted on a stainless steel shaft, supported at each end by glass bearings. The shaft is driven by a synchronous motor and variable gear train, through magnetic coupling. The curvature of the cylinder cross section is high at each end and low in the middle, Fig. 2, which confines the sphere to the central portion of the cylinder, while allowing it to roll in a random manner. A probe, which is in effect one-third of a complete Faraday cage, is mounted concentric with the cylinder and spaced eight millimeters from its inner surface. The probe was calibrated by the use of a cylinder with an insulated segment which, together with the probe, comprised an essentially complete Faraday cage. Comparison of the total charge of the sphere with that measured by the probe alone showed that the probe "collected" 33±4 percent of the charge of the sphere. The probe can be rotated upward until its shielding screens it completely from the charged sphere, where it can be grounded to provide a reference point. The cylinder is kept at ground potential by means of a *graphite brush* pressing against the shaft.

Low capacity, low leakage circuit leads were made from fine wire sealed into glass tubing, subsequently outgassed and evacuated. Shielding is provided by copper foil wrapped around the tubing. Polyethylene and polystyrene cable were found unsatisfactory because of relatively high spurious charges developed by slight flexing. Such charges leak away slowly.

An inverse feedback type of vacuum tube electrometer was developed, in which the potential difference between the measuring circuit and its shielding is essentially zero, thus minimizing charge leakage while allowing the use of relatively high voltage ranges. The effective grid current of the electrometer tube used was about 10^{-15} ampere, which is small compared with the rates of charging encountered in this work. An adjust-

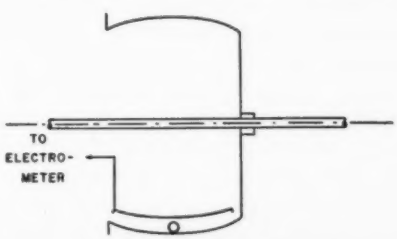


FIG. 2. Cross section of cylinder.

* This work was supported by the U. S. Office of Naval Research.
† Now at Bell Laboratories, Murray Hill, New Jersey.
[1] J. W. Peterson, J. Appl. Phys. **25**, 501 (1954).
[2] D. E. Debeau, Phys. Rev. **66**, 9 (1944).
[3] The spheres were ground to a good polish from carefully annealed material, and were round to within half a wavelength of light. The grinding was done by tumbling the blanks in a cylindrical cavity containing abrasive particles, by means of an air jet.

able air gap condenser of 40 to 255 mmf determines the potential difference produced between the measuring grid of the electrometer tube and the feed-back circuit by the charge induced on the probe. A ten-ohm resistor in series with the electrometer circuit supplied a ten-millivolt full-scale potentiometer type recorder, with a full-scale balancing time of two seconds. The chart drive was equipped with a quick change, allowing paper speeds of six or sixty inches per hour.

The pumping system consisted of a mercury diffusion pump, backed by a high-speed mechanical pump. In addition to the standard trap between the mercury pump and the final stopcock, a second liquid nitrogen-cooled trap prevented stopcock grease from entering the experimental tube. Pressures in the lower range were measured by an ionization gauge, which showed that pressures of $3-5\times10^{-7}$ mm Hg were readily obtained.

The tube could be outgassed at 350°C. The nickel cylinder was cleaned by the use of a jet of very fine abrasive particles suspended in water, to remove impurities introduced in the forming process. Buffing was avoided in order that particles of abrasive not be imbedded in the metal. Final cleaning was by detergent solution, followed by steam cleaning and boiling and subsequent rinsing in distilled water.

It was found that, for the case of quartz and borosilicate glass, washing successively in nitric acid, chromic acid, nitric acid, tap water, and distilled water produced uniformly clean surfaces, as shown by uniform charging characteristics. Other methods, such as steam cleaning and rinsing in acetone, followed by distilled water, left surfaces which were not only not uniform but actually had areas which acquired charge of polarity opposite to that of the clean surface. Evidently very great changes in surface properties are produced by even slight contaminations; such dirty surfaces also display rather high conductivity. It is probable that the contamination was by oils and greases, which undoubtedly produce complex and unpredictable local surface structures.

The cylinder was slightly asymmetric, with the result that the sphere was alternately raised and lowered with respect to the probe as the cylinder revolved. This resulted in a total variation of about seven percent in the charge induced on the probe, producing a "wobble" on the recorder trace. Because of its relative freedom of lateral motion, the sphere traveled somewhat farther than the circumference of the cylinder in one revolution of the cylinder. This effect is important only at low speeds and caused a slight overestimate of the charging rate at these speeds.

The cylinder rotated smoothly under the magnetic coupling except at lowest speeds, where bearing friction occasionally produced a jerking motion. Where oscillation of the sphere developed, the results were discarded. Since probe and cylinder were not concentric in cross section, the proportion of charge collected by the probe varied slightly with displacement of the sphere in a direction parallel to the axis; there was negligible variation for large displacements in the direction of rotation. As the charge of the sphere became high its motion became quite erratic, because of the very large image force attracting it to the cylinder causing it to stick. At a charge of about 4×10^{-9} coulomb, measurement became impossible because the sphere adhered to the cylinder and was at times raised considerably above its normal position before it broke loose and rolled back to the bottom of the cylinder.

In order to estimate the surface conductivity of quartz, the probe was grounded in measuring position, then the electrometer switched to a low range of polarity opposite to that of the charge on the sphere. As charge drained from the sphere to the cylinder, the probe appeared to collect charge of the opposite sign. It is assumed that a nearly uniformly distributed charge produces a given potential distribution over the surface of the sphere; it is also assumed that surface conductivity is uniform over the sphere and is approximately independent of electric field strength. Then, without knowing the form of the potential distribution, it can be expected that the potential and therefore the potential gradient is proportional to the total charge on the sphere. Thus, by knowing the total charge, and measuring the leakage current, the surface conductivity can be found in arbitrary units. This procedure is valid as long as the charge distribution has not departed seriouly from the initial approximately uniform distribution, and holds for the conditions under which the measurement was made in practice.

Because of the relatively high surface conductivity of borosilicate glass, the measured charge on a borosilicate glass sphere decreased rapidly after it had stopped rolling. This allowed an accurate measurement of conductivity directly from the recorder tracing, on the same voltage scale at which the charging curve was made. The decrease in measured charge actually represents principally the redistribution of charge over the surface of the sphere, where charge drains from the upper surface of the sphere down to the region near the point of contact. The actual rate of loss of charge from the sphere appeared to be considerably lower than the probe measurement would indicate. If the sphere were revolved a half turn, after standing for a few minutes, bringing the previous area of contact up near the probe, the probe signal became as large, or larger, than it was when the sphere first stopped rolling.

When a gaseous atmosphere was desired, high-purity nitrogen was used, dried by passing through a long trap cooled by liquid nitrogen. At no time did the charging rate or surface conductivity appear to be changed by the repeated introduction of nitrogen.

Discharging the sphere was accomplished in either one of two ways, depending upon its surface conductivity. Spheres having low surface conductivity, such as fused quartz, were effectively discharged by rolling in an atmosphere of nitrogen of about 4 mm Hg pres-

sure, which is near the pressure for minimum saturation charge, as a result of gaseous discharge between the sphere and the cylinder. While the residual charge after this procedure was not zero, it was relatively small and uniformly distributed. Spheres having relatively high surface conductivity, such as borosilicate glass, were discharged by rolling at low speed, which reduced the residual charge to considerably below that resulting from gaseous discharge. In earlier work, discharging was accomplished by x-ray ionization of the gas in the tube. This required near-atmospheric pressure of gas to be effective and with the equipment available required several hours. For these reasons, as well as to avoid the unknown effects of x-rays upon the properties of the nonconductors, their use was avoided in the present work.

The rate of charging of the sphere was most conveniently measured by the rate of gaseous discharge to the probe. At about 4 mm Hg pressure, gaseous discharge occurred only to the probe, and not to the cylinder. This method measured the total charge collected by the sphere per unit distance rolled and incidentally provided a check on the probe calibration, when compared with the initial slope of the charging curve in vacuum. The agreement was quite good.

EXPERIMENTAL RESULTS

Fused Quartz

Fused quartz acquires a negative charge in contact with nickel. When carefully cleaned, the surface is very uniform, as judged by its charging characteristics. Since the probe was relatively close to the upper surface of the sphere, the charge on the extreme upper part of the sphere was heavily weighted in the induction of charge on the probe. Thus, the probe in effect sampled the surface of the sphere as it rolled underneath it, and if the surface of the sphere were not uniformly charged, a fluctuation was produced in the probe signal. Although uniform charge distribution is the exception rather than the rule, when care was taken to produce such uniformity the signal on the probe varied by one percent or less.

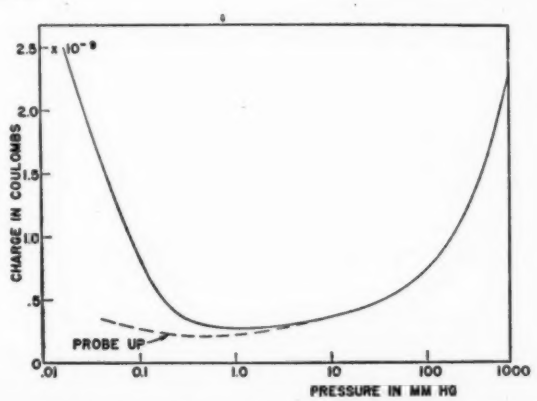


FIG. 3. Maximum charge as a function of pressure of N_2.

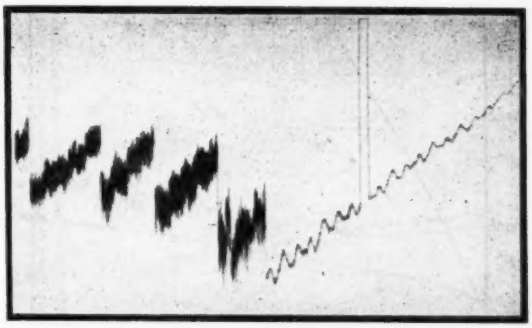


FIG. 4. Photograph of recorder trace, showing large discharges to cylinder; fused quartz, 2.4 cm per second, pressure 400-mm Hg. (Note cylinder "wobble" and point where probe was grounded to check electrometer zero.)

The surface of the quartz sphere became contaminated as a result of outgassing the apparatus, probably by contaminants driven out of the nickel and redeposited on the sphere. Recleaning the quartz invariably returned it to its original charging characteristics. Relatively little contamination appeared to occur as a result of standing in a vacuum of the order of 5×10^{-7} mm Hg. Two quartz spheres from different sources had nearly identical charging characteristics. Therefore, it is probable that results obtained here are generally applicable to fused quartz.

EFFECT OF GAS PRESSURE

The effect of the pressure of the atmosphere in which charging takes place was investigated in order to determine whether gaseous discharge actually is responsible as indicated by earlier work.[1] The maximum charge of the sphere was measured as a function of pressure of N_2, from 3×10^{-7} to 700 mm Hg, giving the results shown in Fig. 3. These results were reproducible to about ten percent and were in principle independent of rolling speed. At low pressures the measured charge was less if the probe had been swung up before measuring the charge than if it had remained in measuring position. The existence of gaseous discharge from the sphere was established by the observation of individual charge transfers through the gas to the probe. Such charge transfer through the gas to the probe can have only occurred by Townsend discharge or spark. Below 10 mm Hg all discharges went to the probe, and above 40 mm Hg all discharges occurred to the cylinder. At higher pressures the discharges were large, provided the rate of acquiring charge was relatively high at this point. Such large discharges resulted in a very nonuniform charge distribution over the sphere, thus producing a large fluctuation in the charge induced on the probe. Figure 4, which is a photograph of the recorder tracing, shows the progress of charging at 400 mm Hg, culminating in several large discharges to the cylinder and the resultant "patchy" charge distribution. Where the charging rate was low, the discharge pulses were often so small even at high pressure as to be indis-

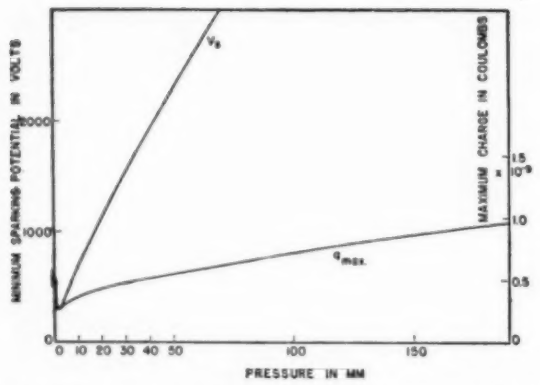


FIG. 5. Comparison of maximum charge of sphere and minimum sparking potential for a pair of electrodes of 5-mm gap (see reference 5), as functions of pressure of N_2.

cernible on the recorder tracing. With decreasing pressure the maximum pulse size decreased, probably tending toward a typical Townsend discharge at very low pressures.[4]

The form of the pressure dependence of maximum charge, the fact that charging rate is independent of pressure, and in particular the fact that gaseous discharges occur to the probe, establish that the pressure dependence of maximum charge is due to gaseous discharge.

The pressure dependence of the maximum attainable charge is very similar to the minimum sparking potential curve for a pair of electrodes, as a function of gas pressure and electrode spacing.[5] To demonstrate the striking similarity, Fig. 5 shows both curves. The usual parameter used in the sparking potential curve is the product pd, where p is the pressure in millimeters of mercury, and d is the gap between the electrodes, in centimeters. It should be observed that, while the minimum sparking potential (V_s) curve is nearly linear in pd above its minimum point, the curve for maximum charge (q_{max}) as a function of pressure falls off gradually, and drops far below the curve for V_s.

While the value of pd for the minimum of the V_s curve is about one in nitrogen, the minimum in the curve for q_{max} is at about two mm Hg. This fact suggests that the discharge occurs from points four or five millimeters from the metal surface. Since the gap between the probe and cylinder is eight millimeters and both are ground potential, the points of maximum potential on the sphere lie at just about this distance. At pressures lower than about one millimeter the probe inhibits discharges, because it effectively limits the path length available for the course of a discharge. With the probe swung up out of the way, however, paths of greater length are available, while the potential of the upper part of the sphere becomes greater, facilitating discharge and lowering q_{max} (see Fig. 3). At approximately one 10^{-3} mm discharges no longer can take place for lack of sufficient potential at any point on the sphere.

As the pressure is increased above that corresponding to minimum charge, discharge becomes possible from points progressively closer to the cylinder. In the hemisphere nearest the cylinder, the potential at a given point of the sphere and the minimum sparking potential at a given pressure both have about the same variation with distance from the cylinder. For this reason, there is a considerable region of the surface of the sphere in which areas of somewhat higher than average charge density can initiate discharges. With increasing pressure, this region moves downward and increases in area. When a discharge is initiated from a point in this region, it may trigger additional discharges until the whole region is discharged. Thus, with increasing pressure the discharges will become larger and leave an increasingly less uniform charge distribution, if the charting rate at this pressure is relatively high, as shown in Fig. 4. If the charging rate is low, there will be greater uniformity of charge distribution, and the discharges may be localized, and therefore small. At pressures below that for minimum q_{max}, a progressively smaller area has sufficient potential to produce a discharge. Therefore, the discharges become smaller, leaving a more uniform charge distribution.

The explanation of the observation that at pressures up to about 10 mm Hg all discharges occur to the probe lies in the probe's greater distance from the sphere. At higher pressures, however, the most favorable distance becomes less than that from the sphere to the probe, and discharges begin occurring to the cylinder.

The largest total charge which can exist on the sphere without producing gaseous discharge will be for the case of uniform charge distribution. After a large discharge

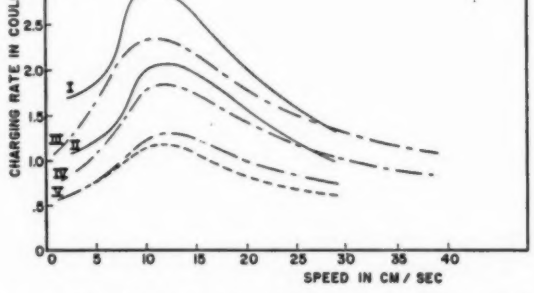


FIG. 6. Charging rate of fused quartz as a function of rolling speed. (I) Freshly cleaned. (II) Freshly cleaned, at 55°C. (III) Contaminated after outgassing apparatus at 225°C. (IV) After outgassing at 325°C. (V) Further contaminated after standing in vacuum for several weeks; dashed portion shows decreased values resulting from long rolling.

[4] It is interesting to note that Townsend discharges occur with glass or quartz cathode and metal anode. This raises the question of γ from such insulators. [W. N. English, Phys. Rev. **74**, 179 (1948), reports discharges from glass in negative point corona.] Here the acquired electrons on the insulator surface may yield a high photoelectric γ, thus facilitating the discharge.

[5] L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (John Wiley and Sons, Inc., New York), pp. 417, 418.

has occurred, draining the charge from a portion of the surface, while the total charge of the sphere has been considerably reduced, other areas of the sphere still have sufficient charge to produce a discharge. Thus, succeeding discharges continue to lower the total charge of the sphere, while producing an extremely "patchy" charge distribution. Finally, a roughly steady state is reached, where the total charge is considerably less than that which could exist if it were uniformly distributed. This effect is shown clearly in Fig. 4.

THE CHARGING RATE

The charging rate, (charge acquired per cm of distance rolled), was measured and found to be independent of pressure in the range from 3×10^{-7} mm Hg to 700 mm Hg. Since charging rate is independent of pressure, it was conveniently and accurately determined by measuring the rate of accumulation of charge on the probe transferred from the sphere by gaseous discharge. Curve I, Fig. 6, shows the charging rate of clean quartz as a function of rolling speed. Curves III, IV, and V show the decreased charging rate due to contamination of the sphere resulting from outgassing the apparatus at 225°C, 325°C, and the further contamination of the latter state resulting from standing for two weeks in a vacuum of about 3×10^{-7} mm Hg. These curves all show a maximum charging rate at an intermediate speed of rolling. Prolonged rolling at higher speeds produced a decreased charging rate, such as that shown by the dashed portion of Curve V. Long rolling at low speed was necessary before the charging rate had recovered its initial value.

Figure 7 gives charging curves for newly cleaned quartz at several speeds of rolling in vacuum, showing the total charge on the sphere as a function of the distance rolled. The curves made at a low speed of rolling, (less than 10 cm per second), are shown beginning at the origin and all have the same general form. The curves for high-speed rolling (greater than 10 cm per second) have been displaced from the origin in the

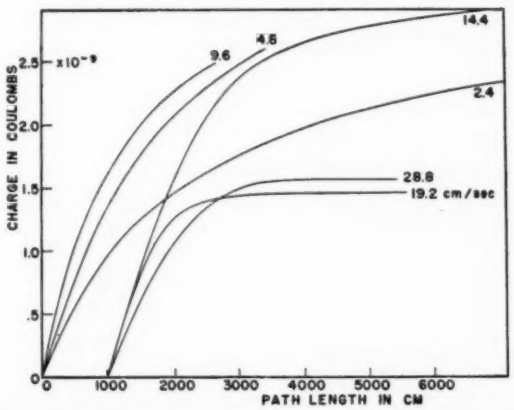


FIG. 7. Charging curves for various rolling speeds, freshly cleaned fused quartz, in vacuum. Curves for speeds greater than ten cm per second have been displaced from the origin for clarity.

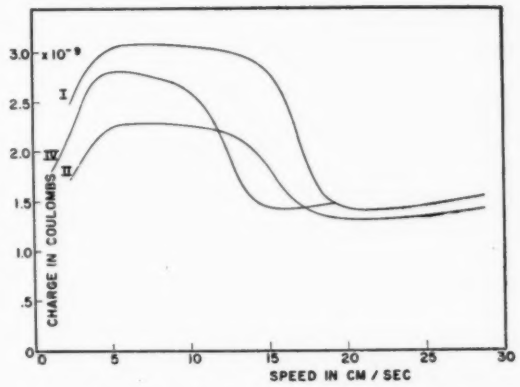


FIG. 8. Maximum charge of fused quartz in vacuum, as a function of rolling speed. (I) Freshly cleaned. (II) Freshly cleaned, at 55°C. (IV) Contaminated after outgassing apparatus at 325°C.

figure in order to show clearly that they all have a similar form different from that for low-speed rolling. The initial slopes of these curves correspond to the charging rates of Curve I, Fig. 6. The maximum charge in vacuum increases with speed up to intermediate speeds and decreases sharply at higher speeds, as shown in Fig. 8. The curves representing the same surface condition in the figures are labeled with the same Roman numeral.

The low maximum charge and altered form of the charging curves for high-speed rolling suggested that a change in surface properties occurs during the course of rolling. Further evidence came from the observation that the maximum charge decreased slightly after continued rolling at high speeds. In order to verify the existence of such a change, the sphere was rolled at 15 cm per second until the maximum charge was reached, stopped for twenty minutes and rolled again at the same speed. The maximum charge increased appreciably after the period of rest, then gradually decreased to the original value. Quite evidently the surface properties after prolonged rolling were different from those after a period of rest.

Such changes could be attributed to heating of the surface, by mechanical working under high electrostatic stresses, with a resulting increase in surface conductivity. To check this hypothesis, measurements were made with the entire apparatus heated to 55°C (30°C above room temperature) using a clean quartz sphere. The charging curves for this case corresponded in form to but lay below those of Fig. 7. The maximum charges, shown by Curve II, Fig. 8, lie considerably below those for clean quartz at room temperature, shown by Curve I. Curve II of Fig. 11 shows the surface conductivity at 55°C, measured immediately after various speeds of rolling. Because of the difficulty of measuring the low surface conductivity of quartz even at this temperature, the curve is only approximately correct. However, it does show a real increase with speed of rolling. Whether conductivity decreases again above 30 cm per second is doubtful, although the curves for

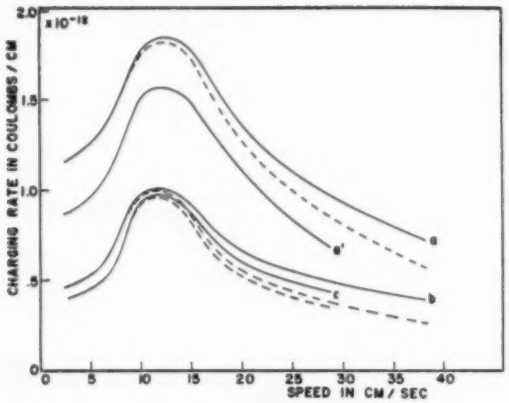


FIG. 9. Charging rate of borosilicate glass as a function of rolling speed. (a) Freshly cleaned (a^1) Contaminated after standing in vacuum for a few days. (b) Further contaminated after outgassing apparatus at 225°C. (c) After outgassing at 325°C. Dashed curves show decreased values resulting from prolonged rolling.

maximum charge all show a slight increase at the high-speed end, as shown by Fig. 8. It is possible that this is caused by minor mechanical instability of motion such as bouncing, though this is not perceptible to the eye.

Surface conductivity can also be measured by the decrease in measured charge of the sphere after it has stood for forty to sixty hours and gives values of conductivity K of about 0.1 at room temperature. Measurements made immediately after the sphere had stopped rolling gave rough values of from 0.4 to 2. Thus surface conductivity appears to increase greatly after rolling, particularly at higher speeds.

Borosilicate Glass

One reason for choosing borosilicate glass was the desire to investigate the charging properties of a material of relatively high surface conductivity, as compared with quartz. Since the surface conductivity of borosilicate glass in vacuum was found to be three to four orders of magnitude higher than that of fused quartz, this condition was fulfilled. Borosilicate glass charges negatively in contact with nickel, and in general, its charging properties are similar to those of quartz.

Considering the difference in conductivity, it is remarkable that the maximum charges and charging rates were of the same order of magnitude as those of quartz. However, borosilicate glass did not seem to have the high degree of reproducibility which quartz exhibited. In addition to becoming contaminated faster and more seriously than quartz, its charging rate at times seemed to decrease after standing for some time with a high charge. This is not surprising, considering that the high local fields are probably sufficient to cause migration of ions, with a resulting change in structure. Migration of ions in such glass has been observed. Such action does not occur with the more stable and strongly bound quartz structures. Subsequent prolonged rolling appeared to cause a recovery of the charging rate reduced by standing with high charge.

The newly cleaned sphere became somewhat contaminated after standing for a day or two in a vacuum of about 5×10^{-7} mm Hg. The surface appeared to stabilize after a few days and remained more or less constant until after the first outgassing of chamber and Ni cylinder at 225°C, after which the contamination had evidently increased, as in the case of quartz. Outgassing at 325°C still further increased the amount of contamination.

Within experimental error the variation of maximum charge with pressure was identical to that of quartz, although higher rolling speeds were required to initiate discharge at the extremes of the pressure range. At the pressure used for charging rate measurements, the discharge pulses to the probe were larger and less frequent than in the case of quartz.

Charging rates are shown in Fig. 9. While having lower values, the curves are quite similar in shape, and the maximum rates of charging occur at about the same speed as for quartz. As in the case of quartz there was a pronounced decrease in charging rates after prolonged rolling at higher speeds, as shown by the dashed curves. Prolonged slow speed rolling returned the charging rates to their original values.

Figure 10 shows maximum charge in vacuum. The curves do not show the sharp decrease at high speeds, characteristic of quartz, indicating that conductivity increases less with speed of rolling than in the case of quartz. This is borne out by Fig. 11, showing the conductivity of both as a function of speed. While the values for quartz are only qualitative, they are accurate enough to establish that the relative increase with speed is much greater for quartz than for borosilicate glass. A comparison of Figs. 10 and 11 shows a definite decrease of maximum charge with increasing conductivity. The speed corresponding to the peak of the maximum charge curve increases with conductivity, as might be expected.

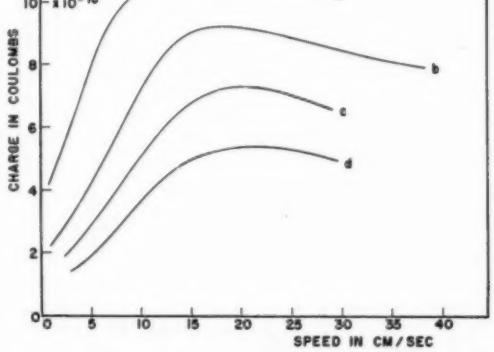


FIG. 10. Maximum charge of borosilicate glass in vacuum, as a function of rolling speed. (a) Freshly cleaned. (b) Contaminated after outgassing apparatus at 225°C. (c) After outgassing at 325°C. (d) Case (c) at 55°C.

INTERPRETATION OF CHARGING DATA

The results just described show that quartz and borosilicate glass have similar charging properties. Charging rate is independent of pressure, but varies with surface contamination. Maximum charge also varies with surface condition, but has a strong dependence upon pressure. Both quantities decrease with increasing surface conductivity, and both show an increase with speed of rolling to a maximum at an intermediate speed. Charging rate decreases sharply for both materials, following this maximum. The maximum charge of borosilicate glass decreases only slightly at higher speeds, while that of quartz decreases sharply and then remains nearly constant. The sign of charge is independent of all these variables, if the sphere has been properly cleaned initially.

Consideration of these results leads to the conclusion that the processes involved in the charging of an insulator in contact with a metal fall rather naturally into two classes: first, that associated with the direction and amount of charge transfer per unit area of intimate contact; and second, those processes which produce the charge distribution which is measured. It seems important to draw a clear distinction between the two, in order to avoid errors in interpretation.

PRIMARY PROCESS OF CHARGE TRANSFER

The primary process of charge transfer is considered to be due to the transfer of electrons,[1] analogous to the electron transfer which produces the contact potential difference between metals of different work function, where the metal of higher work function acquires electrons from the other.

Because the charge transfer is a surface phenomenon, it is impossible to distinguish between a true work function and the existence of surface states. If it is a true work function—that is, a property of the bulk of the nonconductor—it should not be expected to be the same as that measured by other means. For example, a photoelectric work function would of necessity be correlated with filled levels, whereas in the case of a negatively charging nonconductor (as is usually the case) the so-called work function must be correlated with unfilled levels, which can accept electrons. Furthermore, the levels, if they exist, must be localized, or the electrons would be free to move back toward the point of contact. Since the nonconductors investigated have aperiodic structures, it is at present impossible to attempt a calculation of their energy level schemes, and it is perhaps presumptuous even to use the concept of well-defined energy levels or bands. Nevertheless, the reproducibility and the uniformity of charging, under favorable conditions, suggests that there is at least a distribution of levels or states whose average value is well-defined.

There is as yet little direct experimental evidence to support this hypothesis. There is, however, qualitative

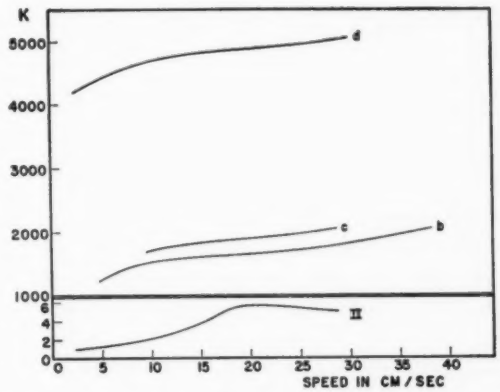


FIG. 11. Surface conductivity as a function of rolling speed. (b) Borosilicate contaminated after outgassing apparatus at 225°C. (c) Borosilicate after outgassing at 325°C. (d) Case (c) at 55°C. (II) Freshly cleaned fused quartz at 55°C.

support from the observation that borosilicate glass charges much less in contact with lightly oxidized nickel than in contact with relatively clean nickel. The electric double layer produced by electronegative oxygen increases the work function of the nickel, thus decreasing the work function difference between it and the glass (which charges negatively in both cases and, therefore, presumably has a higher effective work function).

SECONDARY PROCESSES GOVERNING REDISTRIBUTION OF CHARGE

There appear to be two secondary processes governing the amount and distribution of charge over the surface of the nonconductor: conduction over the surface of the sphere and the gaseous discharge from the sphere where the charge is sufficiently great, which has already been discussed.

Surface Conduction

Assuming an equilibrium contact potential difference to exist between the nonconductor and metal in contact, electrons will be transferred to a region of the nonconductor which is effectively in contact with the metal. The area of the nonconducting sphere in physical contact with the metal can be defined as that area for which the separation between the two surfaces is small enough for electron transfer between the two. This area depends upon the shape and smoothness of the two surfaces, and upon pressure at the point of contact, which can be greatly increased by strong image forces, when the sphere is highly charged. The effective area of contact, to which charge is transferred, is greater than the area of physical contact, however, because of the surface conductivity of the sphere.[6] This follows from the fact that at equilibrium the surface of the sphere would

[6] Volume conductivity is not considered here, since it is generally much lower than surface conductivity, and since, in general, the mobility of the ions to which volume conduction is due is too low, compared with the speed of revolution of the sphere, to produce any important net transfer of charge.

be an equipotential. Thus the degree of equillibrium attained, and therefore the amount and distribution of the charge transferred depend upon surface conductivity and the length of time which an element of area spends near the metal surface. Therefore, it would be expected that this initial rate of *acquiring charge*, (*per unit distance of rolling*), would *decrease* with increased speed of rolling.

As the sphere rolls, the charged area is separated from the metal surface and its potential with respect to the metal increases. Thus a potential gradient is established along the surface of the sphere, producing a flow of electrons back toward the point of contact. The magnitude of this discharging current depends upon the potential gradient and surface conductivity. For a given current, the charge removed from an element of area will depend upon the length of time which it spends in a given high-field region. Therefore, the rate of *loss* of *charge* by *conduction* back toward the point of contact would be expected to *decrease* with increased speed of rolling. The amount of charge removed should not be expected to be inversely proportional to speed of rolling, since the *increased remaining* charge at higher speeds increases the potential gradient and thus the discharging current. In addition, all previously charged areas of the sphere produce a flow of electrons toward the point of contact particularly when the revolution of the sphere brings them close to the metal. This loss again depends on conductivity and time.

Thus, increased speed of rolling tends to decrease the initial rate of acquisition of charge and also to decrease the rate of loss of charge. It is not surprising that the result of these opposing tendencies is to produce a maximum net rate of charging at an intermediate speed of rolling, demonstrated in Figs. 6 and 9.

If it is assumed that, for a given speed of rolling, the rate of acquiring charge is proportional to the uncharged area remaining on the surface of the sphere, and the rate of discharge is proportional to the total charge on the shpere, the *net rate of charging*, (net charge acquired per unit distance of rolling), is given by

$$\frac{dq}{ds}=a\left(1-\frac{q}{q_s}\right)-kq \tag{1}$$

and

$$q=\frac{a}{B}(1-e^{-Bs}), \quad \text{where} \quad B=\frac{a+kq_s}{q_s} \tag{2}$$

$$\left(\text{maximum charge } q_m=\frac{a}{B}\right).$$

q_s represents the total saturation charge of the sphere in the absence of discharging. The quantity $1-q/q_s$ in the early state of charging represents the fraction of the surface of the sphere which has not yet been charged and therefore can accept electrons at the rate a per unit distance rolled. After the sphere has acquired appreciable charge, this quantity also takes account of the fact that previously charged areas have lost some of their charge by leakage, and therefore can accept an additional charge. k is a constant of proportionality for the rate of electron flow from the previously charged areas toward the point of contact and is assumed to be directly proportional to the surface conductivity. Surface conductivity as indicated earlier can be measured. It is represented by the symbol K, which is expressed in arbitrary units, that are equal to the apparent positive current to the probe per volt of probe signal, multiplied by 10^{15}.

By substituting a/q_m for B into (2) yields,

$$q=q_m(1-e^{-as/q_m}) \tag{3}$$

$$\frac{dq}{ds}=a(1-e^{-as/q_m}). \tag{4}$$

All the quantities in these equations are measurable.

Experimentally, as described earlier, it is very difficult to remove all charge from the sphere. It is simple, however, to remove most of the charge, leaving a uniform distribution of relatively low charge density. Therefore, this procedure was used, and the charging curves were made with such an initial uniform charge on the sphere. In this case

$$dq/ds=a(q_s-q_i-q)-k(q_i+q),$$

where q is the increase in charge over the initial value q_i, and

$$q=(q_m-q_i)(1-e^{-as/qB})=q_m'(1-e^{-as/q_m}) \tag{5}$$

$$\frac{dq}{ds}=\frac{(q_m-q_i)}{q_m}(1-e^{-as/q_m})=\frac{q_m'}{q_m}(1-e^{-as/q_m}). \tag{6}$$

Thus, the predicted charging curve for the case of a uniform initial distribution of charge is the same as that for zero initial charge, except for a multiplying factor. The charging rate is smaller by the factor q_m'/q_m. Charging rates measured under these conditions have been corrected to the value for zero charge, in order that they may be compared for different conditions.

To check experimental agreement with Eq. (6), the charging curves for quartz at 55°C, similar to those of Fig. 7, have been replotted on semilogarithmic paper, plotting $(q_m'-q)$ against distance rolled. Equation (5) predicts a straight line for such a plot, and Fig. 12 shows good agreement for the low-speed curves. The similarity of low-speed charging curves for quartz, demonstrated in Fig. 7, is thus explained by the fact that they are all of the predicted form, while the curves for high-speed rolling depart rather seriously. In addition, the maximum charge is much lower at high speeds in the case of quartz.

This effect could be accounted for by either a decrease in the rate of initial electron transfer, or an increase in the rate of discharge, as a result of continued

high-speed rolling. The first possibility seems improbable when it is noted that the net rate of charging is always higher for high-speed rolling than for the lowest speeds, while the low speeds have final charges greatly in excess of those for high-speed rolling. These considerations led to the conclusion that conductivity increases after high-speed rolling, which was subsequently confirmed by the curves of Fig. 11, showing surface conductivity after prolonged rolling at various speeds. The increased surface conductivity of the sphere was attributed to surface heating at high rolling speeds, which seems to be confirmed by the large increase in conductivity measured at 55°C. These results indicate that an increase in surface temperature of about 5°C for borosilicate glass and 15°C in the case of quartz are sufficient to account for the increased conductivity at high speeds. It appears that a relatively shallow surface layer becomes heated, and that it is cooled by conduction to the bulk of the sphere when it is not rolling. This surface cooling appears to take place more rapidly than the uniformly heated sphere itself would cool by conduction, as shown by the rapid recovery of the surface during a period of rest. Because of its much lower thermal conductivity, the quartz surface would be expected to cool less while rolling, and therefore show greater surface heating than borosilicate glass.

It is not surprising that there should be surface heating at high speeds under the rapid application and removal of strong image forces to an element of area, as it approaches and departs from close contact with the metal surface. That relatively high surface charges are required to produce such heating is evident from the comparatively slight decreases in charging rates, after prolonged rolling at 4 mm Hg pressure where gaseous discharge limited the surface charge to low values.

The decrease in the charging rate of quartz due to contamination probably reflects a change in effective work function, as well as an increase in conductivity. This appears to be so, since the rate of charging and maximum charge are less for the contaminated case than for clean quartz at 55°C, although the latter case had the higher conductivity. It is probable that contamination similarly alters the work function of borosilicate glass, while increasing its surface conductivity, as shown in Fig. 11.

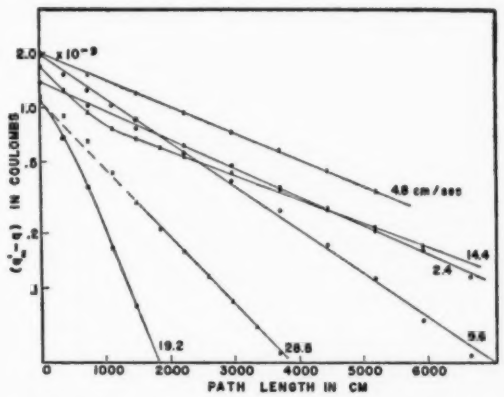


FIG. 12. Semilogarithmic plots of the charging curves for fused quartz at 55°C. Equation (5) predicts a straight line for such plots.

A very rough estimate of the variation of maximum charge with conductivity can be made by comparing, curves *c* and *d* of Figs. 10 and 11. The amount of contamination for these two cases was about the same, and if it can be assumed that a 30°C temperature change produces no important change in the effective work function of rate of electron transfer across the surface of contact, the difference in maximum charge is due to the difference in surface conductivity. If this is true, maximum charge varies roughly as the inverse fourth root of surface conductivity. If the same relation can be assumed to hold for such large differences in conductivity as that between quartz and borosilicate glass, the contact potential difference (and therefore initial electron transfer), between the latter and nickel would have to be at least five times that between quartz and nickel, to account for the fact that the maximum charge for clean borosilicate glass is about half that for newly cleaned quartz.

The author wishes to express his sincere thanks, and to acknowledge his very real debt to Professor Leonard B. Loeb, who suggested this problem and under whose guidance its study was carried out. He would like also to express his appreciation of the many hours spent in difficult glassblowing by Mr. Edward H. Guyon and Mr. Morley R. Corbett, and of the great assistance of Mr. William M. Brower in the construction of many intricate components.

Charge Storage in Junction Diodes*

EARL L. STEELE
Electronics Laboratory, General Electric Company, Syracuse, New York
(Received November 16, 1953)

A calculation is made which relates the physical properties of a junction diode to the troublesome spike current arising from hole storage in the germanium. This spike current results when the diode is suddenly switched from the conducting to the nonconducting state. For small germanium pellets, this current decay is exponential with a decay considerably shorter than that indicated by the hole lifetime in the material. For larger pellets the decay is a sum of exponentials giving a considerably longer decay time, still shorter, however, than the hole lifetime.

1. INTRODUCTION

WHEN a germanium diode is conducting in the forward direction, there exists a greater than equilibrium density of charge carriers in the bulk germanium. Experiments[1] show that, if the diode is suddenly switched from the conducting to the blocking direction, this stored charge manifests itself by producing a sharp "spike" reverse current similar to the sudden discharge of a condenser. At higher frequencies, the size of this spike[2] is large enough to nullify essentially the "uni-directional" conducting characteristics of the diode. The following work relates the physical parameters of the germanium junction diode to the spike current by a calculation based on simple assumptions.

2. ASSUMPTIONS FOR THE ANALYSIS

It is assumed that the diode can be represented as shown in Fig. 1. The problem can then be solved as a one-dimensional case. The piece of germanium (n-type) of thickness (L_0) has a very abrupt diffused p-n junction[3] at one edge ($x=0$), and a metallic surface at the other edge ($x=L_0$). When the diode is conducting in the forward direction, the hole density (p) through the germanium decays from an injected value (p_0) at the junction to a value ($p=0$) at the metallic interface. [Actually, (p) represents the hole density deviation from equilibrium, i.e., when ($p=0$), the hole density is at the thermal equilibrium value.] Also, it is assumed that all the current is carried by holes; the electron contribution is negligible. This can be shown[4] to be the case for diffused junction diodes[5] where the hole conductivity (σ_p) in the acceptor region (p-side of barrier) is much greater than the electron conductivity (σ_n) of the bulk germanium (n-side of barrier). The additional assumption is made that a voltage applied to the diode results in no electric field in the bulk germanium itself; all the voltage drop appears across the p-n junction.[3] Under this assumption, the equation governing the hole density in the germanium is the well-known diffusion equation

$$\partial p/\partial t = -p/\tau + D\partial^2 p/\partial x^2, \tag{1}$$

where (τ) is the lifetime and (D) is the diffusion constant for holes in the germanium.

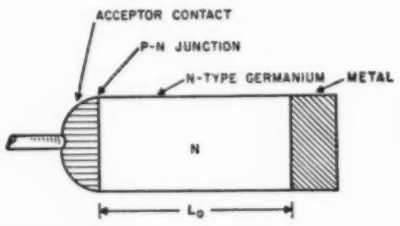


FIG. 1. Schematic diagram of diffused junction diode with abrupt p-n junction.

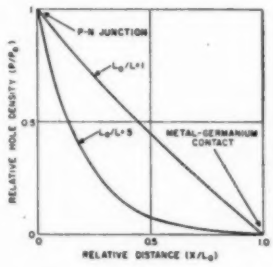


FIG. 2. Hole density distribution through the germanium when a steady state current flows in the forward direction through the diode.

When a voltage (V) is applied to the diode causing it to conduct in the forward direction until a steady state current is reached, the above Eq. (1) can be solved subject to the condition

$$\partial p/\partial t = 0, \tag{2}$$

and to the boundary values

$$p = p_0 \quad \text{at} \quad x=0,$$
$$p = 0 \quad \text{at} \quad x=L_0.$$

This solution for the hole density in the germanium for the steady state current case above is given by Eq. (3).

$$p = p_0\left[\frac{\sinh[(L_0-x)/L]}{\sinh(L_0/L)}\right], \tag{3}$$

where (x) is the distance into the germanium measured from the junction as shown in Fig. 2, and (L) is the well-

* This work has been supported by the USAF Air Research and Development Command, USAF Air Materiel Command, Army Signal Corps, and Navy Bureau of Ships under Contract AF33-(600)17793.

[1] S. E. Michaels and L. A. Meacham, Phys. Rev. **78**, 175 (1950).
[2] R. G. Shulman and M. E. McMahon, J. Appl. Phys. **24**, 1267 (1953).
[3] J. S. Saby and W. C. Dunlap, Jr., Phys. Rev. **90**, 630 (1953).
[4] Earl L. Steele, Proc. Inst. Radio Engrs. **40**, 1424 (1952).
[5] R. N. Hall, Proc. Inst. Radio Engrs. **40**, 1512 (1952).

known "diffusion length," the average distance the carriers travel before recombination

$$L=(D\tau)^{\frac{1}{2}}.$$

The above solution, Eq. (3), defines the non-equilibrium density distribution in the germanium at the moment the diode is switched from the conducting to the blocking direction. It is the discharge of these holes which will cause the spike current.

Let us assume that the diode is conducting in the forward direction due to an applied voltage (V) for ($t<0$). Then at ($t=0$) this voltage is suddenly reduced to zero, which we assume reduces the hole density at ($x=0$) also to zero. The current and hole density as a function of time and distance are then calculated.

3. CALCULATION

The solution for the transient hole density obtained from Eq. (1) must satisfy the hole distribution subject to the following boundary conditions for ($t>0$)

$$p=0 \quad \text{at} \quad x=0,$$
$$p=0 \quad \text{at} \quad x=L_0.$$

This solution is obtained by assuming it to be a product of a function of distance (x) and a function of time (t). The result is the expression

$$p=[c_1\cos(\alpha x)+c_2\sin(\alpha x)]e^{-t/\tau}\exp(-D\alpha^2 t), \quad (4)$$

which when fitted to the above boundary conditions requires

$$c_1=0 \quad \alpha=\alpha_n=n\pi/L_0,$$

where (n) is an integer indicating the root of the equation

$$\sin(\alpha_n L_0)=0.$$

Since the complete solution must fit the initial hole distribution, Eq. (3), for ($t<0$), the complete solution will be a sum of terms like Eq. (4):

$$p=\sum_n c_n\sin(\alpha_n x)\exp[-(t/\tau)(1+\alpha_n^2 L^2)]. \quad (5)$$

The coefficients c_n are then evaluated by the usual Fourier method by requiring Eq. 5 to equal Eq. 3 at $t=0$. The final expression for the transient hole density is then given below:

$$p=p_0\sum_n \frac{2}{n\pi}\sin\left(\frac{n\pi x}{L_0}\right)\left[1+\left(\frac{L_0}{n\pi L}\right)^2\right]^{-1}$$
$$\times\exp\left[-\frac{t}{\tau}\left\{1+\left(\frac{n\pi L}{L_0}\right)^2\right\}\right]. \quad (6)$$

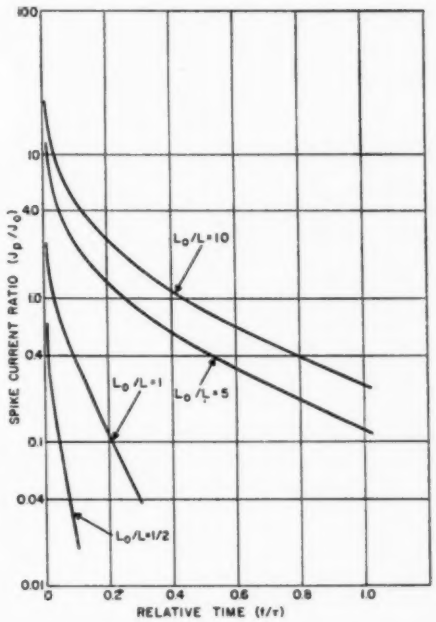


FIG. 3. Semilog plot of spike current *versus* time, illustrating change in decay characteristic for different germanium diode sizes.

The current density is also given by the terms

$$J_p=-eD(\partial p/\partial x)$$

$$J_p=-p_0\left(\frac{2eD}{L_0}\right)\sum_n \cos\left(\frac{n\pi x}{L_0}\right)\left[1+\left(\frac{L_0}{n\pi L}\right)^2\right]^{-1}$$
$$\times\exp\left[-\frac{t}{\tau}\left\{1+\left(\frac{n\pi L}{L_0}\right)^2\right\}\right]. \quad (7)$$

The spike current is obtained by evaluating (J_p) at $x=0$; it is then approximated by the following expressions:

(a) For the case ($L_0/L\ll 1$) only the first term in the sum is important and

$$J_p=(-2ep_0D/L_0)\exp[-(t/\tau)(\pi L/L_0)^2]$$
$$=J_0\exp[-(t/\tau)(\pi L/L_0)^2]. \quad (8)$$

This case yields an interesting result. The "effective time constant" of the diode is reduced from (τ) to $[\tau(L_0/\pi L)^2]$. For instance, when ($L_0/L=\frac{1}{2}$) this decay factor is $(\frac{1}{2}\pi)^2\approx 0.025$, implying a decay constant of about 2.5 microseconds for a diode made from germanium having a hole lifetime of 100 microseconds and a good frequency response to about 400 kc/sec; this also implies a germanium pellet about 10 mils in thickness.

(b) For the case ($L_0/L\gg 1$), several terms enter yielding

$$J_p=J_0\sum_n\left[1+\left(\frac{L_0}{n\pi L}\right)^2\right]^{-1}$$
$$\times\exp\left[-\frac{t}{\tau}\left\{1+\left(\frac{n\pi L}{L_0}\right)^2\right\}\right]. \quad (9)$$

The evaluation of this expression involves considerable numerical work since the number of terms which must be included for accuracy depends upon the value of (L_0/L).

These results have been plotted for several values of the ratio (L_0/L) and are shown in Fig. 3. It can be seen that the sharp gradient introduced at the junction by suddenly reducing the diode voltage to zero causes the holes to surge out, resulting in a large back current, "the spike," which falls off very rapidly as the sharp gradient is decreased. The decay of the current is noted to be extermely rapid for small ratios of (L_0/L).

For longer times the current becomes essentially exponential for large ratios of (L_0/L) but deviates considerably from the simple exponential for $(t/\tau<0.1)$.

The change of hole density distribution with time in the germanium for one value of the ratio $(L_0/L=1)$ is shown in Fig. 4. The sharp gradient mentioned before is obvious from this figure.

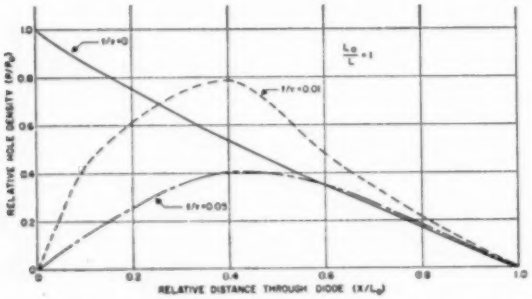


FIG. 4. Graph showing hole density variation through germanium for different time intervals after removal of forward bias voltage on diode. For the case shown, it was assumed that $L_0/L=1$.

4. CONCLUSIONS

The important factors contributing to an improvement in frequency response and a corresponding reduction in spike current appear from the foregoing analysis. A reduction of the lifetime of holes in the bulk germanium as well as the use of thinner pellets will both tend to reduce the spike. The important ratio (L_0/L) is also affected (increased) by a decrease in the lifetime because of the effect on the "diffusion length." This can be compensated for, however, by simply reducing the pellet thickness (L_0).

JOURNAL OF APPLIED PHYSICS VOLUME 25, NUMBER 7 JULY, 1954

Instability of Liquid Surfaces and the Formation of Drops

JOSEPH B. KELLER AND IGNACE KOLODNER
Institute of Mathematical Sciences, New York University, New York, New York
(Received October 19, 1953)

The theory of Taylor instability is extended to take account of surface tension, and the most unstable mode is determined. The resulting theory is then used to estimate the size of drops which might be formed as a result of this instability.

I. INTRODUCTION

IT has long been known that certain flows of liquids with free boundary surfaces are unstable. For steady flows this means that a small perturbation introduced at a point of the flow produces a disturbance which increases in the flow direction until it completely alters the motion. This mechanism was used by Rayleigh to describe the breakup of a steady jet of liquid into drops. In the case of unsteady flows, instability implies that certain initial perturbations will increase in time until they significantly change the flow. Taylor[1] showed that the plane free boundary of a semi-infinite liquid is unstable if the liquid moves with constant acceleration normal to the surface, with the acceleration directed into the liquid. Previously, Penney and Price[2] had shown that the surface of a pulsating underwater explosion bubble is likewise unstable. In this case, also, the fluid moves with accelerated motion normal to the surface.

With these results in mind, we wish to account for the formation of drops resulting from the breakup of accelerated plane, cylindrical and spherical layers of liquid. These phenomena occur whenever explosives are detonated in or near liquids. For example, the spray dome produced by an underwater explosion results from a layer of water thrown into the air by the shock wave from the explosion. The present theory attempts to describe the breakup of this layer into drops of spray. As a second example, consider the production of an aerosol (i.e., dispersion of liquid drops in air) by a bomb containing an explosive surrounded by a layer of liquid. The high pressure gases resulting from the explosion propel the liquid layer outward, and it ultimately breaks up into drops. The present theory also attempts to account for this breakup.

In our proposed mechanism of breakup we assume

[1] G. I. Taylor, Proc. Roy. Soc. (London) **A201**, 192 (1950).
[2] W. G. Penney and A. T. Price, British Report SW-27 (1942).

that there is a zero-order accelerated flow of the liquid layer. The first-order perturbation of this flow, which satisfies linear equations, is represented as a series or integral of normal modes. Some of the modes are found to be unstable in the sense that their amplitudes increase indefinitely with time. Among these, one or more are most unstable, in the sense that they grow most rapidly. In the case of constant acceleration the unstable modes grow exponentially, and the most unstable modes have the largest exponents. The most unstable modes are assumed to be responsible for the breakup of the layer into pieces, which then become spherical due to surface tension. The number of pieces depends upon the number of positive and negative regions into which the most unstable mode function is decomposed by its nodal lines. Thus, it is possible to estimate the number and size of the resulting drops.

In the following section (II) the problem is formulated for the case of a plane layer. In Sec. III the zero-order solution is considered, and in Sec. IV the first-order perturbation is obtained. This perturbation solution is analyzed in Sec. V, where the formation of drops is discussed, and practical conclusions and recommendations are presented. Corresponding results for cylindrical and spherical layers will be given in another paper.

II. FORMULATION

We consider the motion of a plane slab or layer of liquid which separates two gases at different pressures. The liquid will be assumed to be incompressible and inviscid, and the motion of the gases will be neglected. The initial positions and shapes of the bounding surfaces will be prescribed, as well as the initial velocity of the liquid. The problem is to determine the subsequent positions and shapes of these surfaces, and the motion of the liquid.

Assuming that the motion is irrotational, the velocity $u(x,y,z,t)$ of the liquid is derivable from a potential $\phi(x,y,z,t)$ which satisfies Laplace's equation. Thus,

$$\mathbf{u}=-\nabla\phi, \tag{1}$$

$$\nabla^2\phi=0. \tag{2}$$

The pressure $p(x,y,z,t)$ in the liquid can be expressed in terms of ϕ, the constant density ρ, and the acceleration of gravity g by Bernoulli's equation

$$p\rho^{-1}=\phi_t-gz-\tfrac{1}{2}(\nabla\phi)^2. \tag{3}$$

It is assumed that the positive z axis points vertically upward.

The boundary surfaces are assumed to be given by

$$z=F_i(x,y,t) \quad i=1,2. \tag{4}$$

The subscript $i=1$ corresponds to the upper surface. The gas pressure on both surfaces is assumed to be independent of x and y, and is denoted by $p_i(i=1,2)$. Then the dynamic and kinematic boundary conditions may be written, respectively, as

$$\phi_t-gF_i-\tfrac{1}{2}(\nabla\phi)^2-(-1)^i\rho^{-1}T_i2H_i=\rho^{-1}p_i \quad \text{on} \quad z=F_i \quad (i=1,2) \tag{5}$$

$$\phi_z-\phi_xF_{ix}-\phi_yF_{iy}+F_{it}=0 \quad \text{on} \quad z=F_i \quad (i=1,2). \tag{6}$$

In (5), T_i is the surface tension and H_i the mean curvature of the surface i. The initial conditions will be introduced when they are needed.

If the flow is bounded by a cylindrical tube parallel to the z axis, the condition $\partial\phi/\partial n=0$ is imposed on this tube.

If the lower surface $z=F_2$ is a rigid surface with prescribed motion, the above formulation applies provided (5) is omitted for $i=2$. Therefore, this problem will also be considered, since it is so similar to the original problem.

III. ZERO-ORDER SOLUTION

If bounding surfaces are initially planes normal to the z axis, and if the initial velocity is parallel to the z axis and independent of x and y, the solution will be independent of x and y. Then, if $u(t)$ denotes the z component of velocity, we have, from (1) and (2),

$$\phi^0=-u(t)z+b(t). \tag{7}$$

Since F_i^0 depends on t only, we have, from (5) through (7), denoting t differentiation by a dot,

$$\dot{b}=(\dot{u}+g)F_i^0+\tfrac{1}{2}u^2+\rho^{-1}p_i \quad i=1,2 \tag{8}$$

$$u=F_{it}^0 \quad i=1,2. \tag{9}$$

From (9) we have, if h denotes the initial separation between the surfaces,

$$F_2^0=F_1^0-h. \tag{10}$$

From (8), eliminating $\dot{b}$, we obtain, using (10),

$$(\dot{u}+g)\rho h=p_2-p_1. \tag{11}$$

Equation (11) yields $\dot{u}$, and therefore u if its initial value is given, provided p_2-p_1 is constant or a known function of t (or even of u and t). If the pressure difference is constant, $\dot{u}$ is constant. However, if p_2-p_1 depends upon F_2^0, then (11) becomes a second-order equation for $F_2^0(t)$. Finally b is determined from (8), except for an inessential additive constant.

In the problem in which $F_2^0(t)$ is a prescribed moving rigid surface, u and F_1^0 are given in terms of F_1^0 by (9, 10), and $\dot{b}$ by (5) with $i=1$, while (5) for $i=2$ is omitted, as well as its consequence (11).

IV. FIRST-ORDER PERTURBATION

Suppose that the initial boundary surfaces differ slightly from planes, and for that the initial velocity differs slightly from the constant z component assumed in the above solution. Then we may expect the subsequent solution to differ slightly from the above solu-

tion, which we will henceforth call the zero order or unperturbed solution. If ϵ is a measure of the maximum deviations of the initial data, we assume that the solution may be written in the form

$$\phi=\phi^0+\epsilon\phi^1+\cdots, \quad F_i=F_i^0+\epsilon F_i^1+\cdots. \tag{12}$$

From (2) and (12) we see that ϕ^1 also satisfies Laplace's equation. Now inserting (12) into (5) and (6) we obtain for the first-order perturbation ϕ^1 and F^1 the equations

$$\phi_t-(\dot{u}+g)F_i+u\phi_z-(-1)^i\rho^{-1}T_i\nabla^2F_i=0 \quad \text{on} \quad z=F_i^0 \quad (i=1,2) \tag{13}$$

$$\phi_z+F_{it}=0 \quad \text{on} \quad z=F_i^0 \quad (i=1,2). \tag{14}$$

In (13) and (14) and the sequel, the superscript one on ϕ and F is omitted.

To solve Laplace's equation for ϕ we assume that ϕ is a product of a function of z and t multiplied by a function $\psi(x,y)$. We find

$$\phi=[g_1(t)e^{kz}+g_2(t)e^{-kz}]\psi(x,y) \tag{15}$$

$$(\nabla^2+k^2)\psi=0. \tag{16}$$

If the flow is unbounded in the x,y plane, then (16) has a bounded solution for every real value of k, and the general solution for ϕ is an integral with respect to k of the expression on the right of (15), in which g_1 and g_2 depend upon k.

However, if the flow is bounded by a rigid cylinder parallel to the z axis, and if this cylinder intersects the x,y plane in a curve C, then the function ψ must also satisfy the condition

$$\partial\psi/\partial n=0 \quad \text{on} \quad C. \tag{17}$$

If C is closed, then (16) and (17) have solutions only for a denumerable set of positive values of k^2, and for $k^2=0$. In this case the general solution for ϕ is a series of terms of the form given by (15), summed over the various values of k.

In order to complete the solution we must find g_1, g_2, F_1 and F_2. To this end we assume

$$F_i=f_i(t)\psi(x,y) \quad i=1,2. \tag{18}$$

Now, inserting (15) and (18) into (13) and (14) and making use of (16) we obtain

$$\dot{g}_1\exp(kF_1^0)+\dot{g}_2\exp(-kF_i^0)-(\dot{u}+g)f_i+ukg_1\times\exp(kF_i^0)-ukg_2\exp(-kF_i^0)+(-1)^ik^2\rho^{-1}T_if_i=0 \quad (i=1,2) \tag{19}$$

$$kg_1\exp(kF_i^0)-kg_2\exp(-kF_i^0)+\dot{f}_i=0 \quad (i=1,2). \tag{20}$$

From (20) we find g_1 and g_2 in terms of $\dot{f}_1$ and $\dot{f}_2$,

$$g_1=-k^{-1}\exp(-kF_1^0)(\dot{f}_1-e^{-kh}\dot{f}_2)(1-e^{-2kh})^{-1} \tag{21}$$

$$g_2=k^{-1}\exp(kF_1^0)(\dot{f}_1-e^{kh}\dot{f}_2)(1-e^{2kh})^{-1}. \tag{22}$$

Now using (21) (22), and (9) in (19), and combining the two Eqs. (19), we obtain

$$[(\partial^2f_1/\partial t^2)-e^{-kh}(\partial^2f_2/\partial t^2)]+k(\dot{u}+g)(f_1-e^{-kh}f_2)+k^3T_1\rho^{-1}(f_1+T_1^{-1}T_2e^{-kh}f_2)=0 \tag{23}$$

$$[(\partial^2f_1/\partial t^2)-e^{kh}(\partial^2f_2/\partial t_2)]-k(\dot{u}+g)(f_1-e^{kh}f_2)-k^3T_1\rho^{-1}(f_1+T_1^{-1}T_2e^{kh}f_2)=0. \tag{24}$$

Equations (23) and (24) are a pair of linear homogeneous second-order ordinary differential equations for f_1 and f_2. The coefficients are constant only if $\dot{u}$ is constant, which case we shall treat in detail. First, however, we wish to point out the limiting forms which (23, 24) attain when h becomes infinite, namely,

$$(\partial^2f_1/\partial t^2)+k(\dot{u}+g+k^2T_1\rho^{-1})f_1=0, \tag{25}$$

$$(\partial^2f_2/\partial t^2)-k(\dot{u}+g-k^2T_2\rho^{-1})f_2=0. \tag{26}$$

Either of these is the equation for the perturbation of the surface of a semi-infinite region of liquid, in agreement with Taylor's result. We note that, for sufficiently large constant $\dot{u}$, one of the foregoing equations will have a growing exponential solution while the other will not.

If $\dot{u}$ is constant we seek solutions of (23) and (24) of the form

$$f_i=A_ie^{\alpha t} \quad i=1,2. \tag{27}$$

Inserting (27) into (23) and (24), and simplifying, leads to an equation for α and another for A_2/A_1, namely,

$$\alpha^4+\alpha^2k^2\rho^{-1}(T_1+T_2)\coth kh-k^2(\dot{u}+g+k^2T_1\rho^{-1})(\dot{u}+g-k^2T_2\rho^{-1})=0 \tag{28}$$

$$A_2/A_1=e^{kh}(\alpha^2+k[\dot{u}+g]+k^3T_1\rho^{-1})\times(\alpha^2+k[\dot{u}+g]-k^3T_2\rho^{-1})^{-1}. \tag{29}$$

The solutions of (28) for α^2 are given by (when $T_1=T_2$)

$$\rho h^3T_1^{-1}\alpha^2=-x^3\coth x\pm x(\beta^2+x^4\sinh^{-2}x)^{\frac{1}{2}}. \tag{30}$$

The abbreviations used in (30) are

$$x=kh, \quad \beta=\rho h^2T_1^{-1}|\dot{u}+g|. \tag{31}$$

Since there are four values of α, each f_i is a sum of four terms of the form in (27).

In the problem in which $F_2(t)$ is the prescribed motion of a plane rigid surface, $F_2^1(t)$ is zero. All equations through (24) apply, but (13), (18), and (19) hold only for $i=1$ and $f_2=0$ in (21) and (22). Instead of (23) and (24) we obtain for f_1 the equation

$$(\partial^2f_1/\partial t^2)+k\tanh kh(\dot{u}+g+k^2T_1\rho^{-1})f_1=0. \tag{23'}$$

For infinite h this also has the limiting form (25).

V. INSTABILITY AND DROP FORMATION

If α^2 is negative, α is pure imaginary and the corresponding exponential is oscillatory. This occurs when

the negative sign is chosen in (30). Only when the plus sign is chosen is there a range of x for which α^2 is positive, and thus one value of α is real and positive leading to a growing exponential. In this range there is a greatest value of α corresponding to a particular value of x which we will call $x_{\max}$, since the exponential corresponding to this value of x grows most rapidly.

To find $x_{\max}$ we first note that

$$x^4 \sinh^{-2}x < 1.3. \tag{32}$$

Thus, if $\beta \gg (1.3)^{\frac{1}{4}} = 1.14$, (30) becomes (using the plus sign)

$$\rho h^3 T_1^{-1} \alpha^2 = x(\beta - x^2 \coth x). \tag{33}$$

From (28), α^2 is positive for x between zero and $\beta^{\frac{1}{2}}$. It has its maximum at approximately

$$x_{\max} = (\beta/3)^{\frac{1}{2}} = h[\rho|\dot{u}+g|/3T_1]^{\frac{1}{2}}. \tag{34}$$

Recalling that by (31), $x = kh$, we may introduce $k_{\max}$ which is the value of k corresponding to $x_{\max}$, and thus to maximum instability. From (34) we have

$$k_{\max}^2 = \rho|\dot{u}+g|/3T_1. \tag{35}$$

To find the number of pieces into which the layer breaks up due to the growth of the term corresponding to $k_{\max}$, we must examine the function $\psi(x,y)$ corresponding to $k_{\max}$. It is clear that if the layer is unbounded in the x,y plane there will be many functions $\psi(x,y)$ corresponding to $k_{\max}$. On the other hand, if the region is bounded, (17) applies, and there may be no solution ψ corresponding exactly to $k_{\max}$, although there will be many functions ψ corresponding to values of k near $k_{\max}$. Thus, in both cases there will be many functions ψ with exactly or approximately the same value of $k(\approx k_{\max})$, and therefore all the corresponding terms will grow at about the same (maximum) rate. Consequently, the exact manner of breakup will depend upon the extent to which these various terms are excited by the initial perturbation.

Nevertheless, it is possible to make a rough estimate of the number of pieces into which the layer breaks. First, in the case of an unbounded region, one solution ψ corresponding to $k_{\max}$ is

$$\psi = \exp[2\pi i(x/\lambda_1 + y/\lambda_2)], \tag{36}$$

where

$$(2\pi/\lambda_1)^2 + (2\pi/\lambda_2)^2 = k_{\max}^2. \tag{37}$$

We now consider the regions of the x,y plane in which the real part of ψ is positive (or negative). These are rectangles bounded by nodal lines, and their dimensions are $\lambda_1/2$ and $\lambda_2/2$, with area $\lambda_1\lambda_2/4$. If λ_1 and λ_2 are related by (37), the area of such a rectangle is minimized when $\lambda_1 = \lambda_2 = 2\pi\sqrt{2}/k_{\max}$. Thus, the minimum area of such a rectangle is

$$\frac{\lambda_1^2}{4} = \frac{2\pi^2}{k_{\max}^2} = \frac{6\pi^2 T_1}{\rho|\dot{u}+g|}. \tag{38}$$

It is to be expected that (38) will represent roughly the area of a piece into which the layer breaks up since all parts of the surface in this region move in the same direction. The volume of a piece will be given by (38) multiplied by the thickness h. The radius r of the sphere into which the piece will ultimately deform is then [using (11)]

$$r = \left[\frac{9\pi T_1 h}{2\rho|\dot{u}+g|}\right]^{\frac{1}{3}} = \left[\frac{9\pi T_1 h^2}{2|p_2 - p_1|}\right]^{\frac{1}{3}}. \tag{39}$$

The same results are obtained if the layer is bounded by a tube of rectangular cross section, and they will also apply for tubes of almost any other shape.

On the basis of (39) three means are available for diminishing drop size, namely,

(a) Reducing the layer thickness h;
(b) Reducing the surface tension T_1 by using wetting agents;
(c) Increasing the explosive pressure p_2.

Of these the first two seem reasonable, while the last appears impractical. We note from (39) that diminishing h has a twofold effect: the acceleration is greater and the pieces are thinner.

To estimate the time required for breakup to occur, we assume that the surface perturbation, given essentially by (27), reaches the value h. This gives for the break-up time

$$t = \frac{1}{\alpha_{\max}} \log\frac{h}{A} = \left[\frac{27T_1}{4\rho|\dot{u}+g|^3}\right]^{\frac{1}{2}} \log\frac{h}{A}$$

$$= \left[\frac{27T_1\rho^2 h^3}{4|p_2 - p_1|^3}\right]^{\frac{1}{2}} \log\frac{h}{A}. \tag{40}$$

In (40) A is the initial perturbation of the surface, $\alpha_{\max}$ was obtained from (33) and (34), and the last equality follows from (11).

Flicker Noise in Germanium Rectifiers at Very Low- and Audio-Frequencies*

D. K. Baker †
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania
(Received November 3, 1953)

An investigation of the frequency dependence of flicker noise in point contact and diffusion-alloy type germanium rectifiers is reported. The frequency range extends from approximately 10^{-3} to 10^{6} cy per sec and the inverse frequency law is found to hold over this range. The current dependence of the flicker noise at 10 cps is also reported.

I. INTRODUCTION

FLICKER noise power per unit frequency interval is distinguished from other general types of noise (Johnson, Shot) by its dependence on the steady current squared, by its dependence on the inverse of the frequency f and by the fact that it is a very slowly varying function $F(T)$ of the temperature. The flicker noise power $G(f)$ may be represented by a relation of the form,

$$G(f)df = A\frac{F(T)I^2df}{f}, \tag{1}$$

where I is the steady current through the element. $G(f)$ is also known as the spectral density. Recent work[1] has extended the frequency dependence to high frequencies and has shown a change in the spectral density to an f^{-2} frequency dependence at high frequencies. van der Ziel[2] has developed an explanation for this effect.

Since the total noise power generated in an electronic element would be given by the integral of (1) over all frequencies, we see that the f^{-1} relation must change at high frequencies to some other higher value of the index in order that the integral be closed. Likewise, at the low-frequency end the frequency dependence must change to something less than the minus-one power in order for the integral to be closed at the lower limit and for the total noise power to be finite. One part of this paper is an extension of the measurement of flicker noise to very low frequencies in an endeavor to find such a "turnover" at low frequencies.

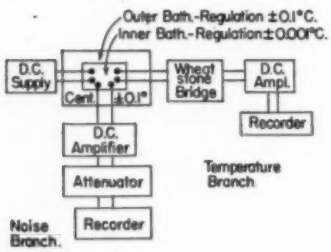


FIG. 1. Block diagram of low-frequency apparatus.

* Supported by the U. S. Office of Naval Research. This material is taken from a thesis submitted to Graduate School, University of Pennsylvania, in partial fulfillment of requirements for the degree of Doctor of Philosophy.

† Now at Physics Dept., Union College, Schenectady, New York.

[1] R. H. Mattson and R. van der Ziel, J. Appl. Phys. **24**, 222 (1953).

[2] A. van der Ziel, J. Appl. Phys. **24**, 1063 (1953).

II. LOW-FREQUENCY MEASUREMENTS

The spectral density $G(f)$ of any fluctuating quantity $y(t)$ which is stationary in time, and whose average value is zero, can be found by the Weiner-Khintchine theorem which relates the spectral density $G(f)$ to the autocorrelation function defined as

$$\langle y(t)\cdot y(t+\tau)\rangle_{Av}, \tag{2}$$

by the relation

$$G(f)=4\int_0^\infty \langle y(t)\cdot y(t+\tau)\rangle_{Av}\cos 2\pi f\tau d\tau. \tag{3}$$

Thus, a determination of the fluctuating (noise) voltage $y(t)$ in a rectifying barrier as a function of time should lead to a determination of the correlation function (2) and the spectral distribution of the noise by Eq. (3).

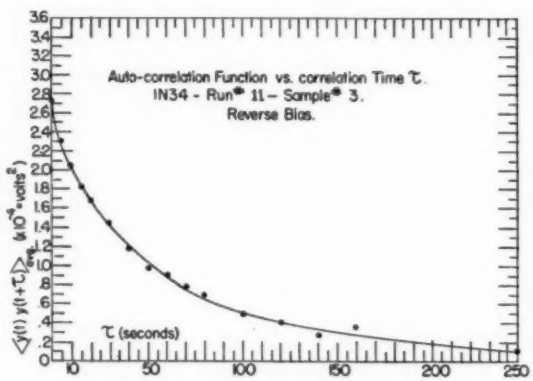


FIG. 2. Auto-correlation function for a point contact crystal rectifier.

Apparatus: Figure 1 is a block diagram for the apparatus showing how the noise voltage from the rectifier, and the temperature of the copper block in which the rectifier was mounted, was brought out to chart recorders. These recorders had a full-scale response time of two seconds and a chart speed of one inch in 10 sec. Thus it is possible to record fluctuations of frequencies up to $\frac{1}{2}$ cy per sec.

It is important to note that measurements were made with the sample in a constant temperature bath. Earlier experiments with low-frequency effects showed a strong correlation between the noise voltage and the ambient temperature of the smaple. This correlation was ob-

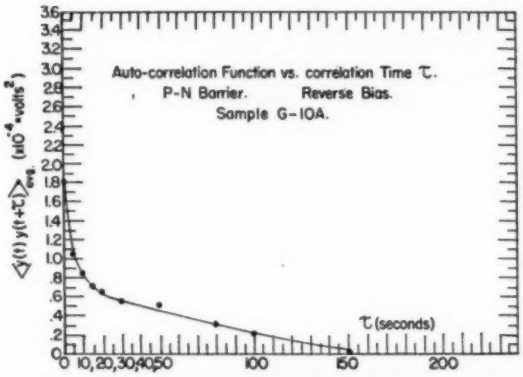


FIG. 3. Auto-correlation function for a diffusion-alloyed type crystal rectifier.

servable even if the temperature was controlled to ±0.01°C. For temperature control better than this, there was no apparent visible correlation.

The rectifier was mounted as one arm of a specially constructed Wheatstone bridge and fluctuations in the current through the rectifier were detected as fluctuations in the null point of the bridge. This bridge and the sample were immersed in an adequately stirred oil bath which gave a constant temperature of 34°C controlled to within ±0.001°C. Temperature measurement was made by means of a rod-type thermistor mounted parallel to, and as close as possible to the sample. Both were mounted in a copper block and immersed in the oil bath. This inner bath was surrounded by two inches of insulating puffed polystyrene and the air region between the inner and outer chamber was stirred by an electric fan and controlled to within ±0.1°C.

Extraneous effects were carefully checked and controlled throughout the experiment. Batteries, which were used for the dc supply were first checked for noise properties when supplying current loads up to ten times greater than normal use. No noise was found. Battery drain and instability was investigated and found to be negligible at currents used in the experiments if the battery circuit was closed 48 hours before measurements were made.

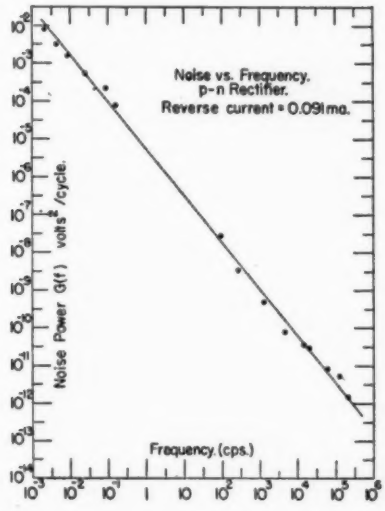


FIG. 5. Flicker noise spectrum—diffusion-alloyed type rectifier.

With the sample mounted, the bath was closed and allowed to stand for at least three or four days. Under these conditions runs could be made for as long as six hours without the temperature changing by more than ±0.001°C. The temperature and the noise records were taken simultaneously on the chart recorders and at the same chart speed so that they could be visually compared. From these records, sections of about 3000 sec long were taken and the correlation function (Eq. (2)) was determined by directly reading the values of y from the curve and forming the products $y(t)\cdot y(t+\tau)$, and calculating average values. The correlation function for a point contact rectifier is shown in Fig. 2 and for a diffused-alloy type rectifier in Fig. 3. The spectral densities were calculated from these curves by evaluating the integral of Eq. (3) numerically with the aid of Simpson's one-third rule. Only that part of the correlation function was used that seemed to give a significant contribution to the integral. The error of omission was determined by fitting an exponential to the correlation function and carrying out the integral. Such errors

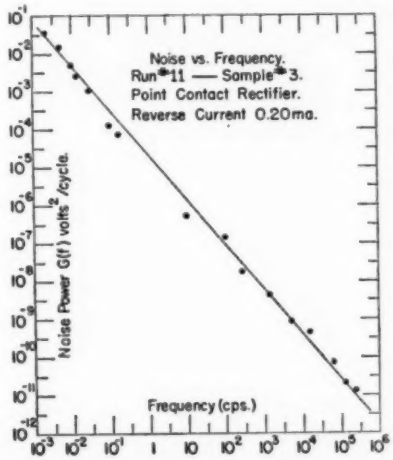


FIG. 4. Flicker noise spectrum—point contact rectifier.

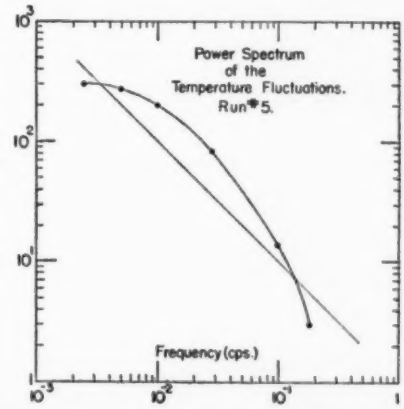


FIG. 6. Power spectrum of the temperature fluctuations.

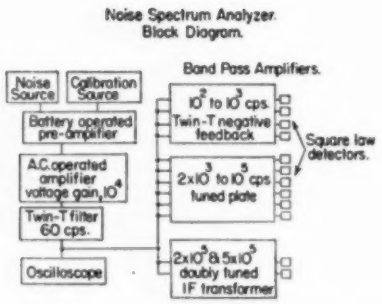


FIG. 7. Block diagram of upper frequency analyser.

were kept below 5 percent. The spectral densities as a function of frequency obtained in this manner are plotted at the low-frequency end of Fig. 4 for the point contact rectifier, and in Fig. 5 for the diffusion rectifier.

To obtain the value of the noise at 10 cy per sec the rectifier was removed from the bath and with the same current through it, the noise was measured by means of an EMR ten-cycle amplifier. The value obtained in this fashion is also plotted in Fig. 4.

The correlation function for the temperature was calculated for one of the runs and its spectral distribution determined. It is shown in Fig. 6, where a line of slope minus one is also drawn for comparison.

III. UPPER FREQUENCY MEASUREMENTS

The measurements were made electronically with the spectrum analyser whose block diagram is shown in Fig. 7. Details of this apparatus may be found elsewhere.[3] The noise signal is applied to a battery operated preamplifier which in turn leads to a broad-band amplifier. Twelve selective amplifiers, peaked at twelve frequencies between 80 cy per sec and 500 kc, then receive the signal and deliver it to 12 square-law detectors. The results of these measurements at a current which was held to the same value as that used in the low-frequency work, is plotted in the upper frequency portion of Figs. 4 and 5.

IV. CURRENT DEPENDENCE

The current dependence of the noise power per unit bandwidth was investigated for each of the rectifiers used for the low-frequency measurements. The characteristic current squared dependence was found for each of the units and a typical curve of one of them is shown in Fig. 8 where the measurement was made at a frequency of 10 cy per sec. It is noted that the slope of the curve begins to change at higher currents through the rectifier. It is suspected that heating effects at the boundary cause changes in the dynamic resistance of the rectifier and lead to a decrease in the noise power at this point.

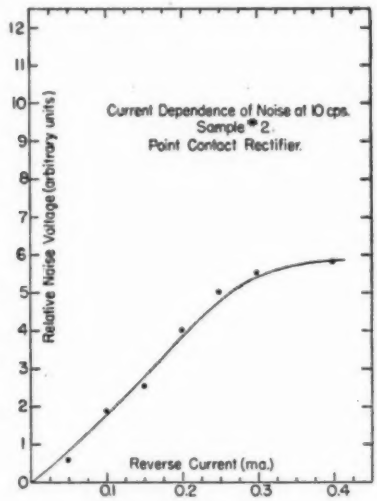


FIG. 8. Current dependence of flicker noise—point contact rectifier.

V. DISCUSSION OF RESULTS

The inverse law of flicker noise has been established for point contact and diffusion alloy-type crystal rectifiers for the frequency range extending from approximately 2×10^{-3} to 10^{6} cy per sec. The measurements were made with currents in the reverse direction and the low-frequency measurements required careful temperature control.

The results showed that there is no "turnover" frequency, as required by the theories, for frequencies as low as 10^{-3} cy/sec. This has the effect of requiring some mechanism, or combination of mechanisms, with an extremely wide distribution of lifetimes for the explanation of flicker noise. No satisfactory explanation of this effect has as yet appeared.

[3] Tech. Report No. 8, Bu Ships cont. No. 34144 University of Pennsylvania.

Letters to the Editor

Low-Loss Ferrites for Applications at 4000 Millicycles per Second

L. G. VAN UITERT, J. P. SCHAFER, AND C. L. HOGAN*
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received April 14, 1954)

SINCE the advent of the microwave gyrator and its associated circuit elements,[1] an extensive search has been made for ferrites which are essentially transparent to microwave radiation when they are subjected to a magnetic field which is just sufficient to take them to magnetic saturation. In the early stages of this search several materials were found which satisfied this condition for frequencies above 7000 mc/sec. However, up to now no ferrites have been reported to have low loss at frequencies of 5000 mc/sec or less.

Ferrites have recently been prepared at Bell Telephone Laboratories, Inc. which are useful at these lower microwave frequencies. The materials are magnesium ferrites modified to lower the magnetic saturation by substituting aluminum for part of the iron. Conductivity losses are suppressed by the aluminum additions and, when these additions are small, by adding manganese and employing less than the stoichiometric amount of iron according to the general formula $MgAl_xMn_yFe_{2-(x+y+s)}O_4^{\pm}$. Faraday rotation and loss curves at 4000 mc/sec for three compositions of this series are shown in Figs. 1, 2, and 3.

A convenient index of the quality of a ferrite for microwave applications is the magnitude of the rotation of the plane of polarization per unit absorption of energy from the transmitted wave. Typical values of this index for the best available ferrites are 400–500 degrees per db at 9000 mc/sec and, until the present work, about 10–20 degrees per db at 4000 mc/sec. A study of the loss mechanisms at microwave frequencies led to the conclusion that the main problem at the lower microwave frequencies was that of obtaining a ferrite with a much lower intrinsic saturation moment than that of the commonly used ferrites, combined with low dielectric loss. This led to ferrites that have given measured factors of merit of greater than 500 degrees per db at 4000 mc/sec.

As normally used for microwave devices, a ferrite element consists of a long slender cylinder supported at the center of a round wave guide. Since the effective lengths of the cylinders used in these experiments are comparable to the guide wavelength, Kittel's[2] equation for determining the applied field necessary for ferromagnetic resonance cannot be applied quantitatively. Nevertheless, for the size and shape of the samples used, it has been found that the equation is certainly valid within several percent. For the case where the length of the cylinder is long compared to the diameter ($1 \geq 10d$) but is still small compared with a wavelength, Kittel's equation predicts that ferromagnetic resonance will occur in an applied field of

$$Ha = (f/2.8) - 2\pi M,$$

where f = frequency in megacycles per second, $2\pi M$ = one-half of the saturation moment of the magnetic material in gauss, and Ha = applied field in oersteds. For radiation with a frequency of 4000 mc/sec, this predicts that the applied field necessary for resonance is given by

$$Ha = 1430 - 2\pi M.$$

In other words, at this frequency a long slender cylinder of ferrite whose saturation moment ($4\pi M$) is about 2860 gauss would be at peak resonance when magnetized even in the presence of a vanishingly small applied field. For reasons explained previously[3] in order to obtain large Faraday rotations with small absorption, the ferrite element must see an effective field which is very small compared to the field required for resonance. Hence for this reason alone it is desirable to use ferrites for microwave devices in which

$$2\pi M < (f/2.8).$$

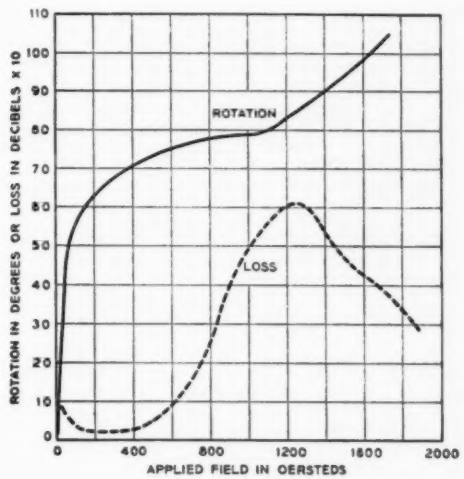


FIG. 2. Rotation and loss as a function of applied field for $4\pi M = 1100$ gauss.

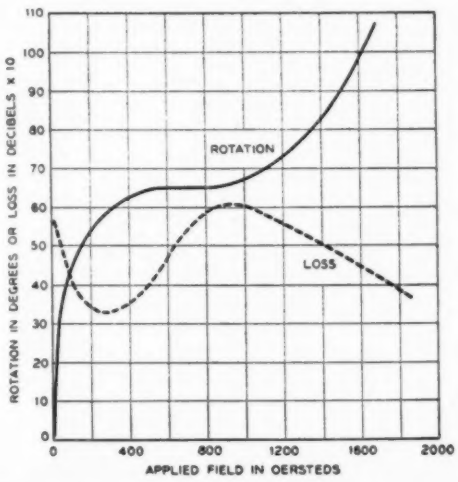


FIG. 1. Rotation and loss as a function of applied field for $4\pi M = 1800$ gauss.

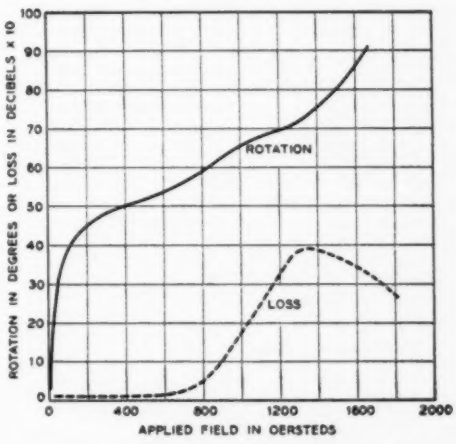


FIG. 3. Rotation and loss as a function of applied field for $4\pi M = 800$ gauss.

For operation at 4000 mc/sec this requires that the saturation moment ($4\pi M$) be certainly less than 2860 gauss. Actually, due to the width of the ferromagnetic resonance absorption line, especially in polycrystalline ferrites, it is necessary that the saturation moment be less than 2000 gauss in order to avoid appreciable absorption due to ferromagnetic resonance even before the sample is completely magnetized.

If this were the only difficulty to be overcome, several of the previously available ferrites would be expected to show low loss at 4000 mc/sec. However, several investigations have demonstrated[4,5] that a completely demagnetized ferrite can show ferromagnetic resonance absorption which will extend up to frequencies as high as

$$f \max = 2.8(4\pi M).$$

The effective magnetic field necessary for resonance is, in this case, supplied by magnetic poles which are created and destroyed on the Bloch walls during the precessional motion of the magnetization vector within each domain.[6] This loss disappears as soon as the sample is completely magnetized but not before. Thus, when a demagnetized sample is placed in a wave guide and the insertion loss is measured as a function of the applied magnetic field, the general shape of the curve is similar to those shown in Figs. 1 and 2. As in Fig. 2, there is normally a strong absorption which decreases as the applied field is increased. The loss is then approximately independent of field and is believed to be primarily dielectric loss. The observed loss then increases again as ferromagnetic resonance is approached. If, as explained above however, the saturation moment is as high as 2000 gauss, these two loss peaks will actually meet (Fig. 1). In order that there should be a useful working range of applied field between these two loss peaks (Fig. 2), it is necessary that the saturation moment be considerably less than 2000 gauss. If the initial low field loss is to be completely absent at 4000 mc/sec (Fig. 3), then the intrinsic saturation moment must be less than 1430 gauss.

The pertinent data concerning the loss peaks in Figs. 1, 2, and 3 are summarized in Table I. It will be noted from the graphs that

TABLE I

Fig.	Ferrite cylinder dimensions	Measured $4\pi M$	Ferromagnetic loss peak Ha	$\Delta\frac{4\pi M}{2}$	Ferrite[a] composition
1	0.38 in. ×3.00 in.	1800	950	350	$MgMn_{0.1}Fe_{1.8}$
2	0.46 in. ×3.00 in.	1100	1250	150	$MgAl_{0.1}Mn_{0.1}Fe_{1.8}$
3	0.46 in. ×3.00 in.	800	1350		$MgAl_{0.3}Fe_{1.7}$

[a] Materials fired at 1350°C for 10 hours.

the insertion loss behaves in essentially the manner to be expected from the foregoing theoretical considerations. In Fig. 2, the low field loss is still present even though the measured saturation moment is less than the 1430 gauss intrinsic value at which it should disappear. This occurs because the intrinsic saturation of the ferromagnetic particles in the ferrite is considerably higher (>1500 gauss) than the measured value of 1100 gauss since optimum density has not been realized.

Kittel's equation for determining the position of the ferromagnetic resonance loss peak cannot be applied quantitatively. However, as shown in Table I, the trend is satisfactorily predicted by the equation. One can expect by a straightforward application of Kittel's equation that the change in the position of the peak will be equal numerically to one-half of the change in the saturation moment. Comparison of the applied fields at which the loss peaks occur with the differences in $2\pi M$ in Table I indicates that the agreement is quite satisfactory.

* Present address: Room 301 Pierce Hall, Harvard University, Cambridge, Massachusetts.

[1] C. L. Hogan, Bell System Tech. J. 31, 1–31 (1952).
[2] C. Kittel, Phys. Rev. 73, 155–161 (1948).
[3] C. L. Hogan, Revs. Modern Phys. 25, 253, 263 (1953).
[4] Rado, Wright, and Emerson, Phys. Rev. 80, 273 (1950).
[5] D. Polder and J. Smit, Revs. Modern Phys. 25, 89 (1953).

A Continuously Recording Electron Diffraction Camera for Studies of Crystal Structure Transitions

A. BOETTCHER, R. THUN, AND H. TREUPEL
Metall-Laboratorium der Deutschen Gold- und Silberscheideanstalt, Hanau, Germany
(Received March 12, 1954)

THE electron diffraction method has proved to be particularly valuable for studying the structure of surfaces, surface layers, and thin films. Another important application of electron diffraction permitted by the short exposure time of electron diffraction patterns is the recording of rapid structure changes in solids caused by transformation and diffusion phenomena. It is obvious that the study of such phase transitions promises substantial aid in the interpretation of diffusion, corrosion, and transformation processes. The present article describes a new electron diffraction camera[1,2] which permits the continuous recording of both structure transformations and the distribution of different crystalline components in a flat specimen.

Figure 1 shows a schematic drawing of the apparatus. The outstanding new features of the apparatus are a viewing screen with a slit diaphragm and a photographic plate which moves with constant speed behind the screen in a direction perpendicular to the slit. If a powder diffraction pattern is formed by transmission of fast electrons through a thin polycrystalline specimen, the direct electron beam and two small portions of each diffraction ring pass through the slit of the screen and expose the moving photographic plate. The resulting pattern consists, therefore, of parallel lines instead of rings. From the distances of corresponding lines, the d values of the investigated material and its structure can be determined.

The specimens can be heated electrically to study phase transitions as functions of temperature and time. The specimens can also be moved synchronously with the photoplate to determine the distribution of various crystalline components over an extended flat sample.

The efficiency of the apparatus is characterized by its time resolution and its geometrical resolving power. The time resolution

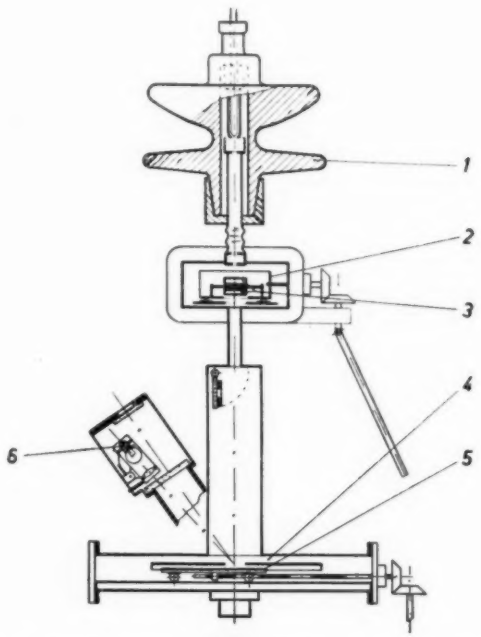


FIG. 1. Electron diffraction camera for continuous recording. (1) Electron gun assembly, (2) specimen holder, (3) specimen heater, (4) viewing screen with slit diaphragm, (5) plate carriage, (6) viewing window and optical arrangement for making light marks.

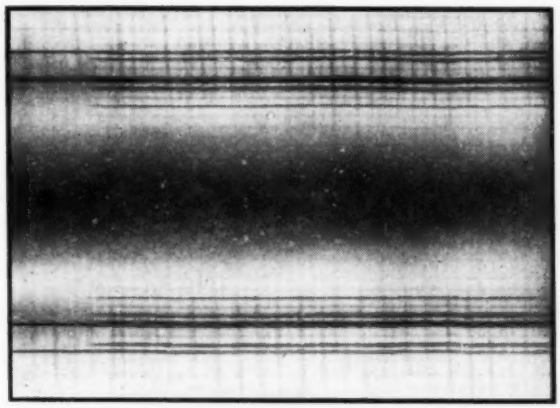


FIG. 2. Recording of the lattice transformation of Ag_2S.

is equal to the exposure time required to take a photograph of the electron diffraction pattern. It is approximately one second. This means that a transition state that exists for less than one second cannot be recorded. The geometrical resolving power is determined by the diameter of the electron beam in the specimen plane. In the apparatus shown in Fig. 1 this diameter is 50 microns. Both resolving powers can be improved easily by the application of electron lenses.

Results:—In pursuance of former investigations,[3] studies were made of the diffusion processes in thin evaporated films of two metals, one deposited on top of the other. For example, the formation of the various intermetallic phases or compounds formed between Ag and Al as well as Cu and Sb was recorded in the temperature range from 20° to 700°C. In addition, several crystal structure transformations were investigated and the changes at the transition temperatures recorded. Figure 2 shows the recorded transformation of Ag_2S which takes place at 180°C. Below the transition temperature, Ag_2S crystallizes in the face-centered orthorhombic system, while a body-centered cubic lattice is stable at higher temperatures. An exact investigation of the structure of Ag_2S as function of temperature shows that the original orthorhombic phase first changes slowly into a tetragonal structure which is then converted rapidly into the cubic high temperature modification of Ag_2S. The formation of the tetragonal phase may be interpreted as an increase of lattice symmetry due to anisotropic expansion of the lattice with increasing temperature, rather than a real structure transformation. Many other transformations of materials which, like Ag_2S, consist of more than one modification were investigated with the apparatus described. Most of the transformations recorded, however, are more complicated and complex than that of Ag_2S. A more complete paper describing studies on structure transformation with the new continuously recording diffraction camera will be published soon.

[1] A. Boettcher, German Patent D. P. No. 845 275, September 7, 1949.
[2] A. Boettcher and R. Thun, Physik. Verhandl. 3, 115 (1952).
[3] A. Boettcher, Z. angew. Phys. 2, 193 (1950).

A Note on the Oxide Replica in Electron Microscopy

SHIGETO YAMAGUCHI
Scientific Research Institute, Ltd., Hongo, Tokyo, Japan
(Received March 29, 1954)

THE purpose of the present study is to show that we can find out the orientation of a metal single crystal *without etching* by the oxide replica method.

The test piece, a Permalloy (Ni, 81.5 percent) single crystal was cut from an ingot with a jeweler's saw.[1] Care was taken to prevent straining of the crystal by excessive pressure or heat. The smooth surface thus prepared was readily oxidized in a fused bath of sodium and potassium nitrates.[2] The oxide film freed from the substrate in an ethanol-bromine solution was magnified in an electron microscope. In the micrograph obtained the octahedral figures regularly oriented are recognizable, as is seen in Fig. 1.

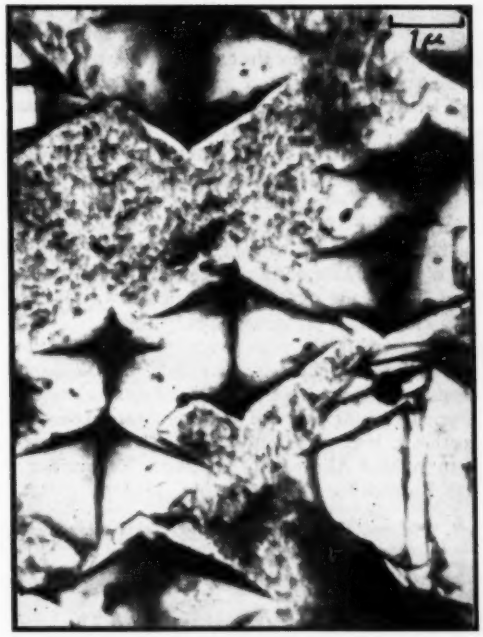


FIG. 1. Micrograph of the oxide film obtained from the nonetched surface of Permalloy cut with special caution. This figure informs us not only of the orientation of the substrate but also of the satisfactory cutting procedure without detectable injury to the substrate.

FIG. 2. Micrograph of the oxide film obtained from a deeply etched surface of Permalloy. The octahedral figures found in this figure are similar to those in Fig. 1.

The finished surface of Permalloy single crystal was rather deeply etched with an ethanol-bromine solution. The oxide film formed on and freed from the substrate gave a micrograph of Fig. 2. The octahedral figures characteristic of Fig. 2 are similar to those in Fig. 1. This means that the oxide film obtainable from a nonetched surface mechanically cut with special caution is able to inform us of the orientation of the substrate. The cutting was so carefully performed here that deep destruction of the surface layer of the single crystal was satisfactorily avoided. This is realized from observation of Fig. 1.

There is found a noticeable difference existing between the thickness of oxide films, corresponding to the octahedral edges in Fig. 1 and that in Fig. 2. This phenomenon was already pointed out and discussed in a previous paper.[3]

It was perceived in the present study that the oxide replica of *etched* surface of Permalloy is more liable to be marred by the presence of rough grains of oxide than the surface *mechanically cut*.

The Permalloy single crystal used here was prepared by Dr. R. M. Bozorth and his co-workers and given to the author with permission to publish the results obtained. The author wishes to thank Dr. Bozorth for his cooperation.

[1] Walker, Williams, and Bozorth, Rev. Sci. Instr. **20**, 947 (1949).
[2] See E. M. Mahla and N. A. Nielsen, J. Appl. Phys. **19**, 378 (1948); S. Yamaguchi, J. Appl. Phys. **23**, 445 (1952).
[3] S. Yamaguchi, J. Appl. Phys. **23**, 935 (1952).

Second Mode Vibrations of the Pochhammer-Chree Frequency Equation*

C. W. CURTIS
Lehigh University, Bethlehem, Pennsylvania
(Received March 19, 1954)

IN a recent communication appearing in *Nature*,[1] Bishop notes that, although the Pochhammer-Chree frequency equation provides an infinite number of modes for describing the propagation of longitudinal elastic waves in a cylindrical bar, experimental studies of strain pulses[2] have failed to reveal contributions from any mode other than the first. Recently, evidence has been obtained for the existence of second mode vibrations.

This evidence appears in records of the elastic strain produced in a magnesium bar, 12 ft long and 1.5 inches in diameter, by suddenly subjecting one end to a uniform pressure, i.e., to a load which is a step function in time. To accomplish the loading, one end of the bar is used to terminate the channel of a conventional shock tube. The pressure is applied as the air shock generated in the tube reflects from the end of the bar. Measurements giving the longitudinal strain as a function of time have been obtained at various distances from the stressed end of the bar by resistance gauges cemented to the side. If it were not for dispersion,[2] the time variation of the strain at all positions along the bar would be the same as the time variation of the force applied to the end. Actually, the strain-time curves resemble a step function only in rough outline. (See Fig. 1.) After propagating a few diameters along the bar, the head of the pulse travels with the "bar velocity," i.e., the velocity determined by Young's modulus. With the arrival of the head, the strain increases rapidly but not as abruptly as the applied force, and there follow oscillations whose beginning amplitude is of the order of 10 percent of the initial rise. Near the head, the oscillations are quite regular and have the appearance of a damped sine wave with a gradually decreasing period. Shortly, however, this regularity is interrupted by the superposition of a second set of much higher-frequency oscillations. Note the behavior near the times marked *S* in Fig. 1. The oscillations then become quite complex but later resolve themselves again into simple vibrations with a slowly varying period. Finally there is an abrupt decrease in amplitude, which appears above the mark *F* in Fig. 1(a). The intervals between the head and onset of the high-frequency vibrations and between the head and abrupt decrease in amplitude both increase in direct proportion to the distance the pulse has traveled, indicating that these groups have definite velocities which are less than the bar velocity. In Fig. 1(b), the abrupt decrease in amplitude occurs at a later time than shown on the record.

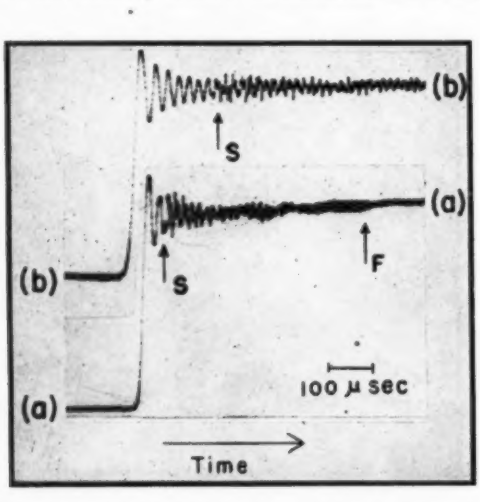


FIG. 1. Longitudinal strains produced in cylindrical bar by reflection of air shock from end. Records (a) and (b) differ in the distance the pulse has traveled from the stressed end of the bar. (a) 150 cm travel. (b) 557 cm travel (reflection from free end).

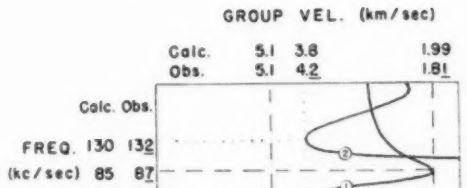


FIG. 2. Comparison of observed strains with predictions based on group velocity analysis. (1) First mode. (2) Second mode.

These observations are in reasonable agreement with the approximate group velocity analysis illustrated in Fig. 2. The curves in this figure give values of the group velocity, calculated from the Pochhammer-Chree frequency equation[3] for the first and second modes, as a function of the frequency about which the groups are centered. The point of particular interest is that the onset of the high-frequency vibrations travels with a velocity in approximate agreement with the calculated value for the fastest group of the second mode and that the frequency of these vibrations corresponds to the predicted value. This is interpreted to mean that these high-frequency vibrations result from contributions of the second mode.

The strain records used in this study were obtained by Mr. George Fox and Mr. Walter Scott. Thanks are due Professor C. A. Shook for advice concerning the mathematics involved in the calculations.

* Work supported in part by the Office of Ordnance Research, U. S. Army.
[1] R. E. D. Bishop, Nature **172**, 169 (1953).
[2] R. M. Davies, Trans. Roy. Soc. (London) **A240**, 375 (1948).
[3] A value of 0.35 was used for Poisson's ratio.

A Note on the Elastic Wedge Problem

J. H. Huth
Pacific Palisades, California
(Received January 18, 1954)

CONSIDER an infinite wedge with an opening of α radians, and introduce polar coordinates r and θ with origin at the apex. Normal and shearing stresses are specified on the two faces $\theta=0$ and $\theta=\alpha$. The problem of determining the stress field within the wedge has been discussed by Timoshenko and Goodier,[1] and their solution appears in series form. However, for each term, the evaluation of the coefficients requires the solution of four simultaneous equations. This can be somewhat simplified through the use of complex variable theory.

Setting the problem to be one of plane strain, the stress components can be written in terms of two analytic functions of the complex variable $z=re^{i\theta}$ as follows[2]:

$$\begin{aligned}\sigma_\theta-i\tau_{r\theta}&=\omega(z)+\bar{\omega}(z)+\frac{z}{\bar{z}}[z\bar{\omega}'(z)+\bar{\lambda}(z)],\\ \sigma_\theta+\sigma_r&=2[\omega(z)+\bar{\omega}(z)].\end{aligned}\tag{1}$$

Assume that the boundary conditions may be expressed as

$$\begin{aligned}\sigma_\theta-i\tau_{r\theta}&=\sum_n (a_n-ib_n)r^n, \quad \theta=0\\ \sigma_\theta-i\tau_{r\theta}&=\sum_n (c_n-id_n)r^n, \quad \theta=\alpha.\end{aligned}\tag{2}$$

One can map the wedge into a half-plane with the function $w=z^{\pi/\alpha}$. Equations (1) and (2) then become

$$\begin{aligned}&\sum_n (a_n-ib_n)w^{\alpha n/\pi}=\omega(w)+\bar{\omega}(w)+\frac{\pi}{\alpha}\bar{w}\bar{\omega}'(w)+\bar{\lambda}(w), \quad \theta=0\\ &\sum_n (c_n-id_n)(-w)^{\alpha n/\pi}=\omega(w)+\bar{\omega}(w)+\frac{\pi}{\alpha}\bar{w}\bar{\omega}'(w)\\ &\qquad\qquad +e^{-2i\alpha}\bar{\lambda}(w), \quad \theta=\alpha.\end{aligned}\tag{3}$$

The problem can now be split into two parts; first one and then the other face will be taken as stress free.

If the face $\theta=0$ is stress free ($a_n=b_n=0$), then ω can be extended to the lower half of the w plane through the relation

$$\lambda(w)=-\omega(w)-\bar{\omega}(\bar{w})-\frac{\pi}{\alpha}w\omega'(w).\tag{4}$$

This reduces the problem to finding a function ω analytic in the entire w plane with the exception of a branch line running along the real axis from the origin to minus infinity. Substitution of Eq. (4) into the second of Eqs. (3) yields the "cut" condition along the axis

$$\begin{aligned}\sum_n (c_n-id_n)(-w)^{\alpha n/\pi}&=\bar{\omega}^+(1-e^{-2i\alpha})+\omega^+\\ &\quad -e^{-2i\alpha}\omega^-+(1-e^{-2i\alpha})\frac{\pi}{\alpha}\bar{w}\bar{\omega}'^+.\end{aligned}\tag{5}$$

A + (plus sign) superscript indicates that the axis is to be approached from above; a − (minus sign) superscript, from below.

This functional equation is easily solved by taking ω in the form $(C_n+iD_n)(-w)^{\alpha n/\pi}$. The result is a set of two simultaneous equations for C_n and D_n in terms of c_n and d_n. When $n=0$, one must take ω in the form $(C_0+iD_0)\ln(-w)$ plus a constant.

Once ω has been determined in this fashion, then λ follows from Eq. (4) and the stress field from Eqs. (1).

If the face $\theta=\alpha$ is stress free we extend ω to the lower half-plane by means of the expression

$$\lambda(w)=-e^{-2i\alpha}\left[\omega+\bar{\omega}(\bar{w})+\frac{\pi}{\alpha}w\omega'(w)\right].\tag{6}$$

This leads to the "cut" equation

$$\sum_n (a_n-ib_n)(w)^{\alpha n/\pi}=\bar{\omega}^+(1-e^{2i\alpha})+\bar{\omega}^+-e^{2i\alpha}\omega^-+(1-e^{2i\alpha})\frac{\pi}{\alpha}\bar{w}\bar{\omega}'^+$$

along the positive half of the real axis. A solution is now obtained by taking ω in the form $(w)^{\alpha n/\pi}$.

Effectively this approach replaces Goodier and Timoshenko's four simultaneous equations by two sets of two equations each. While only a slight simplification has resulted, the method is interesting and holds promise of being applicable to much more difficult problems.

If the wedge is finite rather than infinite and an arbitrary loading is prescribed on the outer boundary, then one would need to also look for nontrivial solutions of either Eqs. (5) or (6) with a_n, b_n, c_n, and d_n set to zero. This leads to eigenequations of the type discussed by Williams.[3]

[1] S. Timoshenko and J. N. Goodier, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1951), p. 123.
[2] N. I. Muskhelishvili, *Singular Integral Equations* (Commonwealth of Australia, Department of Supply and Development, 1949), p. 285.
[3] M. L. Williams, J. Appl. Mechanics **74**, 526 (1952).

A Note on "Effect of Filament Voltage on the Plate Current of a Diode"

Henry S. C. Chen
Drexel Institute of Technology, Philadelphia, Pennsylvania
(Received March 25, 1954)

IN a paper published under the above title in the *Journal of Applied Physics* [**23**, 1254–1256 (1952)], Mr. Ivey discussed the effect of the potential drop across the filament on the plate current of a diode. The effective plate potential is taken as a function of the position along the filament. Using the $V^{\frac{1}{2}}$ law and integrating over the entire filament, he obtained expressions for the factor g which is the ratio of the actual current to that which would flow for equipotential filament.

We would like to point out that for the case when emission is collected from only a portion of the filament, the distortion of the flow lines complicates matters. In the equation for space-charge limited diode, the current is proportional, directly to $V^{\frac{1}{2}}$, but inversely to d^2, where d is the filament-anode distance. Distortion of flow lines requires that d as well as V should be treated as functions of the position along the filament.

To obtain the variation of d, some simplifying assumption has to be made. Following Ivey's notation, if $x+w<1$, then the effective emitting fraction of the filament, corresponding to AB in Fig. 1, is $x+w$. Consider the cathode and the anode, both of length L, separated by a distance d. We shall assume that emission from a point P on the filament shall be collected at Q on the anode, such that

$$\frac{zL}{(x+w)L}=\frac{kL}{L}$$

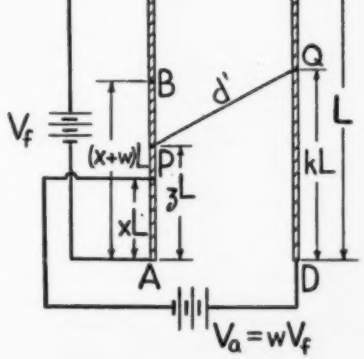


FIG. 1. Geometry of the electrodes.

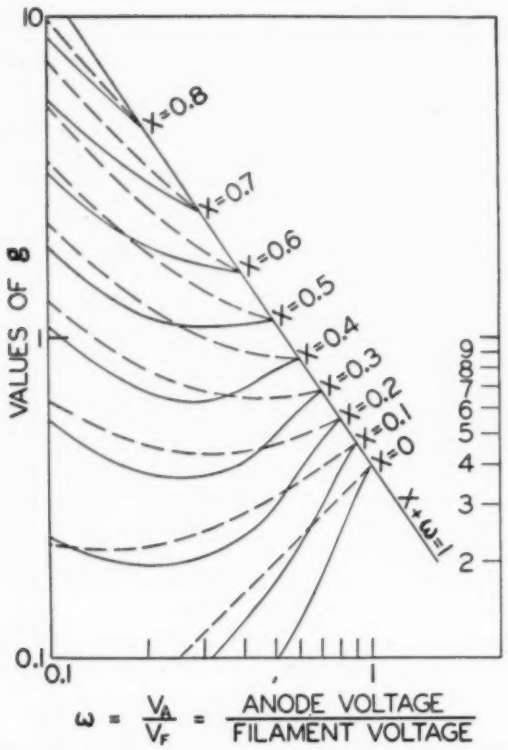


FIG. 2. Values of g as a function of the ratio of anode voltage to filament voltage. The writer's values are in solid curves, while Ivey's values are in dashed curves.

or

$$KL=\frac{z}{x+w}L.$$

Thus

$$d'^2=d^2\left[1+\frac{L^2}{d^2}\left(\frac{1-w-x}{w+x}\right)^2z^2\right]. \tag{1}$$

Therefore

$$g=\int_0^{w+x}\frac{\left(1+\frac{x-z}{w}\right)^{\frac{1}{2}}dz}{1+\frac{L^2}{d^2}\left(\frac{1-w-x}{w+x}\right)^2z^2}. \tag{2}$$

Putting $a=1/K=d/L[(w+x)/(1-w-x)]$ and integrating, we get

$$\begin{aligned}g=a^2\left(\frac{w+x}{w}\right)^{\frac{1}{2}}\Bigg[&K\tan^{-1}k(w+x)-\frac{3}{2(w+x)^{\frac{1}{2}}}\ln\frac{a^2+(w+x)^2}{a^2}\\&+\frac{3\cdot1}{2^2\cdot L^2}\frac{1}{(w+x)^2}\left\{(w+x)-a\tan^{-1}k(w+x)\right\}\\&+\frac{3\cdot1\cdot1}{2^3\cdot L^3}\frac{1}{(w+x)^3}\left\{\frac{(w+x)^2}{2}-\frac{a^2}{2}\ln\frac{a^2+(w+x)^2}{a^2}\right\}\\&+\frac{3\cdot1\cdot1\cdot3}{2^4\cdot L^4}\frac{1}{(w+x)^4}\left\{\frac{(w+x)^3}{3}-a^2(w+x)+a^3\tan^{-1}k(w+x)\right\}\\&+\frac{3\cdot1\cdot1\cdot3\cdot5}{2^5\cdot L^5}\frac{1}{(w+x)^5}\left\{\frac{(w+x)^4}{4}-\frac{a^2(w+x)^2}{2}\right.\\&\left.+\frac{a^4}{2}\ln\frac{a^2+(w+x)^2}{a^2}\right\}+\cdots\Bigg].\end{aligned} \tag{3}$$

Calculations are made for the ratio $d/L=1/10$. Values of g are shown in solid curves in Fig. 2, and are seen to be consistently lower than Ivey's values. The correct values probably lie somewhere between the two solutions. It is interesting to note that for $x=0.5$, g is still greater than unity for all values of w, as in Ivey's solution.

The Resistance of a Bar with a Transverse Hole

W. R. SMYTHE
California Institute of Technology, Pasadena, California
(Received March 25, 1954)

IN collecting material for the *American Institute of Physics Handbook*, we were unable to find a formula for the increase of resistance of a rectangular bar when a circular hole is drilled through it. This is a common laboratory problem so it may be well to put on record a simple and accurate formula which is easily derived from Richmond's conformal transformation.[1] A circular hole of radius c is drilled on the center line of a bar of width a, thickness b, and resistivity τ. The following formula gives an upper limit to the increase of resistance of the bar produced by the hole. For a small hole far from the ends it is practically exact. When the diameter of the hole equals half the bar width, the error in ∇R is still of the order of only 0.1 percent if the distance of the center of the hole from the nearest end equals or exceeds $3a$. The formula is

$$\nabla R=-(2\tau c/ab\theta)\ln\cos\theta,$$

where θ is a parameter chosen so that

$$\sin\theta=\tanh\pi c\theta/(a\theta-\pi c).$$

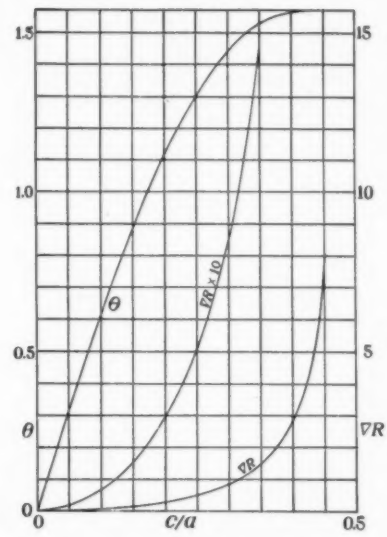


FIG. 1.

When $c=0$ and $\theta=0$, then $\nabla R=0$ and when $c=\frac{1}{2}a$ and $\theta=\frac{1}{2}\pi$, then $\nabla R=\infty$ because the bar is cut off by the hole. For c less than $0.25a$, the formulas will give much more accurate values than can be read from the graph (Fig. 1) which shows θ, ∇R, and $\nabla R\times10$ as a function of c/a. A picture of the flow lines appears in Richmond's paper.

[1] H. W. Richmond, Proc. London Math. Soc. Ser. 2, **22**, 389 (1924). W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill Book Company, Inc., New York, 1950), p. 99.

Erratum: On a Nonlinear Diffusion Equation Applied to the Magnetization of Saturable Reactors

SHOU-HSIEN CHOW
Research Center, Burroughs Corporation, Paoli, Pennsylvania
[J. Appl. Phys. **25**, 377 (1954)]

THE second line following Eq. (18b) should read

$$-(H_0-H_s)/x_s, \text{ etc.}$$